Title
Gel-Seq: An Approach for Simultaneously Sequencing the Genome and Transcriptome in Small Populations of Cells

Author
Hoople, Gordon Donald

Publication Date
2016

Peer reviewed|Thesis/dissertation

Gel-Seq: An Approach for Simultaneously Sequencing the Genome and Transcriptome in Small Populations of Cells

by

Gordon Donald Hoople

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Mechanical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Albert P. Pisano, Co-chair
Professor Liwei Lin, Co-chair
Professor Lisa Pruitt
Professor David Winickoff

Spring 2016

Gel-Seq: An Approach for Simultaneously Sequencing the Genome and Transcriptome in Small Populations of Cells

Abstract

Gel-Seq: An Approach for Simultaneously Sequencing the Genome and Transcriptome in Small Populations of Cells

by

Gordon Donald Hoople

Doctor of Philosophy in Engineering - Mechanical Engineering

University of California, Berkeley

Professor Albert P. Pisano, Co-chair

Professor Liwei Lin, Co-chair

The advent of next generation sequencing (NGS) has fundamentally changed genetics research. Where researchers were once focused on sequencing the genome of a species, they now can sequence the genome of a particular tumor or even a single cell. NGS has also made it cost effective to sequence the RNA transcripts found within a cell, a collection of data known as the transcriptome. Unfortunately, current protocols for sequencing the genome and the transcriptome are incompatible. This leaves researchers with a choice: for a given sample you can examine either the DNA or the RNA. The work presented here makes it so that researchers no longer have to make this choice. This dissertation describes the development of a new protocol, known as Gel-Seq, that makes it possible to sequence both DNA and RNA from as few as 400 cells. This technology will allow researchers to directly examine the ways that changes in the genome impact the transcriptome. At the heart of the Gel-Seq protocol is the physical separation of DNA from RNA. This separation is achieved electrophoretically using a newly designed combination of polyacrylamide membranes that take advantage of the size differences between these molecules.

Two different device options were developed as a part of the Gel-Seq protocol. One device, designed for rapid adoption, was fabricated using standard equipment found inside of a biology laboratory. The second device, designed for separating low input samples, was fabricated using newly developed micro-scale fabrication techniques. In addition to the development of these physical devices, a biological protocol was developed to generate genome and transcriptome data using these devices. In order to validate this technology, a cell line with a stable genome and transcriptome was used. Comparing the Gel-Seq protocol to standard protocols, the results showed a high correlation for both the genome ($R = 0.88$) and transcriptome ($R = 0.96$) data. This supports the conclusion that the device can be used to produce correlated genome and transcriptome libraries. This dissertation reports on the development, optimization, and validation of the Gel-Seq protocol.

To my wife

Contents

List of Figures

List of Tables

Acknowledgments

Let me begin by saying I could double the length of this dissertation with a thorough acknowledgement of all the people that made this work possible. The countless faculty, staff, and students I have worked with over the past 5 years at both UC Berkeley and UC San Diego have been an integral to my success. With that said, there are a few people in particular I would like to acknowledge.

Thanks first to my advisor, Albert P. Pisano, for his incredible support. Not only did he guide me through this process, but he built (and paid for!) the lab of collaborators that made this work possible. Thanks also to my co-advisor Liwei Lin, who, in addition to his technical insights, has helped me navigate my nebulous status as a researcher at both Berkeley and San Diego. Thanks to Kun Zhang for welcoming me into his group and providing a voice of reason on all things bioengineering. Thanks to Andrew Richards for helping to craft this project, without him I wouldn't know the difference between a genome and transcriptome. Thanks to David Rolfe and Kristen Dorsey, who, having traveled this path before, helped show me the way. Thanks to the other members of both Pisano and Zhang labs, all of whom I begged, cajoled, or otherwise berated into helping me with various parts of this project. Finally thanks to my family. Without my parents, who instilled an insatiable desire to learn, or my wife, who pushed me to keep going, none of this would have been possible.

Chapter 1

Introduction

Almost every cell in the human body contains a copy of the entire human genome, encoded using deoxyribonucleic acid (DNA). What makes each cell behave differently is the way in which this genetic code is interpreted. The first step in interpreting this code is to transcribe DNA into messenger ribonucleic acid (mRNA). All of the mRNA molecules, or transcripts, in a cell are collectively known as the transcriptome. By analyzing either the genome and transcriptome researchers are able to the way genetics influences cell behavior.

One major challenge with genetics research is that protocols for analyzing DNA and RNA are incompatible. Researchers are forced to choose between analyzing either the genome or the transcriptome for a particular sample. If samples are sufficiently large they can be split in half. In this case half the sample can be used to find the genome and half to find the transcriptome. The problem with this approach is that large samples tend to average out interesting variations between cells. Researchers are increasingly interested in investigating the variations present in small populations of cells. This, taken to the extreme, has lead to the creation of an entire body of research examining the genomes and transcriptomes of single cells.

In order to motivate the importance of studying small cell populations, consider researchers studying cancer. Tumors are composed of multiple cell types with a range of variations in both DNA and RNA. In order to understand the heterogeneity present in a tumor, it would be useful to profile small groups of cells from different locations. When collecting just a few hundred cells, splitting the sample in half is not feasible. Currently this leaves researchers with a choice, they can either examine the genome or the transcriptome for that sample. This creates a fundamental problem, where researchers are unable to make a direct link between how changes in DNA impact RNA expression. This dissertation presents a solution to this problem

This dissertation describes Gel-Seq, a new protocol that makes it possible to sequence both the genome and transcriptome from a a few hundred cells. At the heart of this protocol is a device that physically separates DNA from RNA. By separating these nucleic acids, and then recovering them into separate tubes, it becomes possible to simultaneously analyze the genome and transcriptome.

This chapter provides the background necessary to understand the Gel-Seq protocol. The first section provides a brief biology primer for those unfamiliar with genetics. This is followed by a section describing the state of the art in genetic sequencing. Next alternative approaches sequencing the genome and transcriptome discussed. Finally, in the last section, an outline for the remainder of the dissertation is provided for the reader.

1.1 Biology Primer: DNA and RNA Inside the Cell

The human genome is roughly 3 billion base pairs long. Nearly every cell within the human body contains a complete copy this genome, though there are slight variations between cells. These variations, known as genetic mosaicism, are very rare and are being investigated for connections to diseases ranging from cancer to Alzheimer's disease [1, 2, 3].

DNA is a long polymer chain composed of four molecules: adenine (A), thymine (T), guanine (G), and cytosine (C). These molecules combine to form a double stranded helix, where each molecule, or nucleotide, is paired with its complement (A-T, C-G). The strands of DNA are composed of sub-units of information known as exons and introns. When combined, exons form larger units called genes that contain instructions for building particular proteins. Introns, on the other hand, exist primarily as DNA and are not expressed in fully formed mRNA.

DNA is stored within the nucleus of a cell in tightly compacted structures known as chromosomes. Humans have 23 pairs of chromosomes, 22 autosomes and 2 sex chromosomes. Each chromosome is made up of strands of DNA with different lengths, the size distribution is shown in Figure 1.1. These lengths are measured in base pairs, the total number of A-T and C-G units in the polymer. The total mass of DNA found within human cells is roughly 6 picograms.

While the DNA contained within each cell is nearly identical, RNA is differentially expressed. Liver cells have a different RNA profile than heart cells. While DNA is designed to store information across generations, RNA is meant to be a temporary messenger. The temporary nature of RNA places an additional time constraint on collecting information about the transcriptome [4]. Ideally samples should be processed within hours of collection to accurately capture the RNA profile of a cell before it degrades. It is also possible to prevent RNA degradation by flash freezing samples with liquid nitrogen. These samples can be safely stored at -80°C for several weeks, however there will be some degradation of the sample.

RNA is made from the same base molecules as DNA, however thymine has been replaced with uracil (U). Additionally RNA occurs only in a single stranded configuration, therefore length is measured in nucleotides. There are three major categories of RNA contained within a cell, messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA) [5]. mRNA is transcribed from the genes found in DNA and moves this information from the nucleus to the cytoplasm where it can be converted to protein. When investigating the links between the genome and transcriptome, researchers are most interested in changes in how

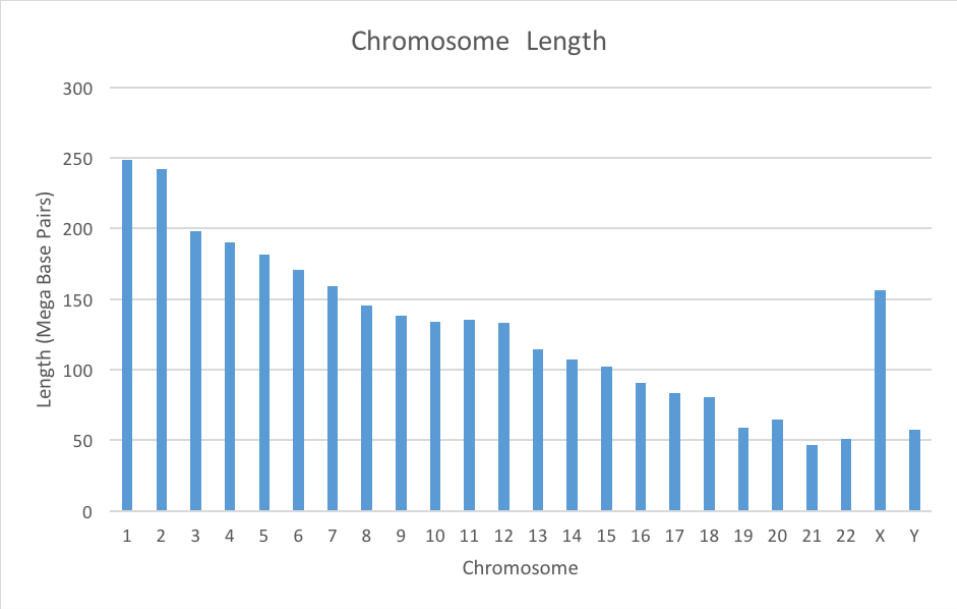


Figure 1.1: The length of chromosomes in the human genome in megabase pairs. Data compiled from the Human Genome Project Rev GRCh38.p6 [7].

variations in DNA impact the expression of mRNA. rRNA and tRNA, both found in the ribosome, facilitate the translation of mRNA into proteins.

While RNA expression varies widely between cell types, a single human cell contains on the order of 1-30 picograms of RNA. Of this total RNA, only 1-10% is the mRNA of interest [5]. The length of RNA contained within the cell is orders of magnitude smaller than chromosomal DNA. While the shortest chromosome is roughly $50 * 10^6$ base pairs long, the longest RNA transcripts are only 10,000 nucleotides long. The majority of RNA transcripts contained within human cells range from 500 - 4,000 nucleotides in length [6].

When building a tool for examining the connections between the genome and transcriptome, there are a range of areas on which to focus. One application of interest for DNA is to examine copy number variations. In a healthy cell there should be two copies of every gene, however research has shown that in certain cells there are additional or reduced copies of sections of the genome. This is known to be a driver of diseases like cancer and may also be responsible for Alzheimer's disease [1, 2, 3]. For examining the transcriptome, generally the most important metric is the number of RNA transcripts present in a cell. Comparing these transcript counts across different types of cells makes it possible to see which genes are being differentially expressed. By collecting information about both copy number variations and transcript counts, it makes it possible to examine how extra copies of DNA impact the expression of RNA.

Figure 1.2: The steps required to obtain a genetic sequence from a biological sample.

1.2 State of the Art in Genetic Sequencing

The National Human Genome Research Institute announced in 2003 they had successfully sequenced the entire human genome. Since then both the cost and technology for genetic sequencing has undergone a dramatic transformation. In 2014 it became possible to sequence a complete human genome for $1000, a remarkable achievement when compared to the $100,000 cost in 2003 [8, 9]. This rapid reduction in cost is due in large part to the introduction of a new generation of sequencing tools, a technology classified broadly under the heading next generation sequencing (NGS).

The steps for obtaining the genetic sequence for a given sample follow the workflow shown in figure 1.2. The first step, sample preparation, relates to the isolation of DNA or RNA. There are a wide range of approaches to extracting and isolating DNA or RNA from cells in preparation for sequencing [10]. Regardless of the approach, the cells must be lysed open to expose the DNA and RNA. This is usually accomplished using a combination of salts and detergent. Generally the DNA or RNA is then isolated from therest of the lysate, however this is not required in all protocols.

In the case of RNA, there is an additional step required during sample preparation. RNA must be converted into complementary DNA (cDNA), a process known as reverse transcription. This critical step is important for a number of reasons. First, DNA is more stable than RNA and is less likely to degrade. Second the rest of the sequencing protocols, as well as many other biological protocols, have been designed around manipulating DNA molecules. This ability to convert RNA to cDNA makes much of the research into RNA possible. This requirement to convert RNA to cDNA, however, is what forces researchers to choose between analyzing the genome or the transcriptome for a given sample. Once the RNA has been converted into cDNA it becomes difficult to distinguish between genomic DNA and cDNA.

The second step in the workflow is amplification. When working with small populations of cells, it is often necessary to create additional copies of the DNA and RNA before it can be sequenced. Typically sequencing requires the equivalent of DNA from 150 cells or mRNA from 1000 cells [11]. The challenge with amplification is to ensure uniform amplification across the genome or transcriptome. Researchers are interested in examining changes due to biological variations, not due to biases introduced from non-uniform amplification. Therefore

extensive research and development time has been invested in developing methods for the uniform amplification of DNA and RNA.

For DNA, a range of approaches have been developed for whole genome amplification [12]. Amongst these, multiple displacement amplification (MDA) has emerged as the current standard of amplification [13, 14]. MDA is an isothermal amplification protocol that leverages the strand displacement activity of the enzyme ϕ29. Single stranded DNA is primed using random primers, typically six bases long, that provide relatively uniform coverage across the genome. Starting from these primers, multiple ϕ29 enzymes polymerize new DNA fragments up to 10 kb in length. When an enzyme encounters a region of DNA that has already been processed into double stranded DNA, it simply displaces this strand and continues the polymerization process. Random primers then anneal to this newly displaced strand and another ϕ29 enzyme binds to the strand and begin the polymerization process again. This process repeats multiple times, creating a branched structure of amplified DNA. While MDA can be susceptible to bias, Gole et al. have shown that performing MDA reactions in microwell arrays can help to reduce this amplification bias [14].

Multiple approaches have also been developed for amplifying RNA [15, 16, 17]. While there are some differences, the methods can be broadly summarized as follows. The first step is to convert the mRNA into cDNA. mRNA has a unique attribute in that the end of the molecule is a string of adenines, known as the poly-a tail. Reverse transcription requires a primer template, so to selectively isolate mRNA a poly-t primer is annealed to the poly-a tail of the mRNA molecule. An enzyme known as a reverse transcriptase then binds to the hybrid and converts the remaining mRNA to cDNA. Once converted cDNA, there are several strategies that can be used to perform uniform amplification. These strategies all involve adding known sequences to each end of the cDNA molecule to amplify using the polymerase chain reaction (PCR).

Once there is sufficient input material, it is possible to proceed to the library preparation step. Library preparation refers to the process converting the amplified DNA into a structure that can be read by the next generation sequencing machines. This entails creating fragments of DNA between 200 and 800 base pairs that have special sequences, known as adapters, attached to either end of the fragment. Different sequencing technologies require different types of library preparation. The sequencing platform used for the work in this dissertation, Illumina, is the current industry leader. Libraries can be prepared for use on Illumina sequencers in a single day from 1 ng of starting material using the NexteraXT kit [11].

Once libraries have been prepared, they can be loaded into the sequencing machine. The sequencer used in this dissertation was the Illumina MiSeq. While a full accounting of the operation of this tool is beyond the scope of this dissertation, the basic operation is as follows. The prepared DNA or cDNA fragments are introduced into a fluidic channel and bind to DNA primers attached to the surface of the channel. The bound DNA is then clonally amplified and spatially constrained by the surface bound primers. Once sufficient amplification has taken place, sequencing occurs by monitoring the incorporation of fluorescent nucleotides onto a complementary DNA sequence [18, 19].

After the genetic sequences of the fragments have been read by the sequencer, the final

step is to computationally assemble the data. With human samples, this is achieved by aligning each fragment to a unique location on a reference genome or transcriptome. (There are other approaches for de novo sequencing when there is no reference available.) Once the fragments have been aligned, the data can then be analyzed in different ways depending on the particular application. Research into both the data processing and analysis aspects of genomic sequencing is a large research topic within the field of Bioinformatics.

1.3 Alternative Approaches for Sequencing Both DNA and RNA

The ability to amplify and sequence either DNA or RNA from small starting samples has only been achieved in the last five years [14, 15, 17]. Consequently there has been very little work regarding how to sequence both DNA and RNA from the same sample. To date there are only two publications on this topic, both from 2015 and both having taken a very different approaches from the work presented in this dissertation.

Dey et al. have developed a new protocol, dubbed DR-Seq, for simultaneously amplifying and sequencing DNA and RNA from the same single cell [20]. DR-Seq relies on a pre-amplification step where both DNA and cDNA derived from RNA are amplified in the same tube. The sample is then split in two and is further amplified before being sequenced. In order to distinguish between genomic DNA and the cDNA derived from RNA, DR-Seq takes a computational approach. Sequences where only exons are present are computationally suppressed, as those could have originated from either DNA or RNA. Instead data for the genome is determined using data based only sequences containing introns. The major drawback of this approach is that it requires a priori knowledge (exons vs. introns) of the genome that is not always known.

Macaulay et al. have developed G&T-Seq, a method for separating, amplifying, and sequencing DNA and RNA from the same single cell [21]. This approach relies on a physical separation of RNA from genomic DNA. This separation is achieved by capturing mRNA on a magnetic bead using a biotinylated oligo-dT primer. Once the mRNA has been captured the beads are held in place with a magnet and the supernatant containing the genomic DNA can be removed and transferred to another tube. Once this physical separation is complete, separate libraries can be generated from the mRNA and DNA. This approach is excellent for working with single cells, however it requires expensive reagents and careful attention to a detailed protocol. For applications where researcher can start with a few hundred cells, the approach taken in this dissertation will be both cheaper and more efficient.

1.4 Gel-Seq Protocol Overview

The fundamental problem of sequencing both DNA and RNA can be solved by taking advantage of the vast size differences between these molecules. To understand these size

differences, consider how DNA and RNA are imaged. DNA exists on the micron-scale and can be viewed using traditional microscopes, while RNA exists on the nanometer scale and must be imaged using complex techniques such as cryo-electron microscopy [22]. The approach taken in the Gel-Seq protocol leverages this size difference to create a physical separation between DNA and RNA.

The basic approach to separating DNA and RNA is shown in Figure 1.3. Panel A shows DNA and RNA free floating in solution near a membrane. When an electric field is applied, as shown in panel B, DNA and RNA experience an electrophoretic force that induces migration through the membrane. (See Chapter 2 for more about electrophoresis.) By tuning the membrane properties, it is possible to create a semi-permeable membrane that separates DNA from RNA. The DNA molecules are pushed against the membrane, but become trapped at the edge because of their large size. Small RNA molecules, on the other hand, are able to reconfigure and weave their way through the semi-permeable membrane much like a snake through grass. Eventually these RNA molecules are then stopped by a second, high density membrane. Once they have been physically separated, the DNA and RNA can be recovered and processed to generate information about the genome and transcriptome.

As was mentioned previously in Section 1.1, RNA is designed to be a temporary messenger and is subject to degradation. Therefore when separating DNA and RNA, it is of paramount importance to ensure the RNA is not damaged or destroyed by the process. Two approaches have been developed as a part of Gel-Seq to ensure RNA integrity during separation. The first approach focuses on stopping the degradation process. RNA degradation is caused primarily by enzymes, known as RNAses, that are present in every cell and digest RNA. By working quickly, maintaining sterile conditions, and adding RNAse inhibitors it is possible to maintain RNA integrity during separation.

A second approach to solving this problem is to perform reverse transcription (RT) before separation. In this case, the unstable RNA molecules are converted into stable cDNA molecules. After RT the cDNA molecules are the same length as the RNA template, so a size based approach to separation remains viable. Recall from Section 1.1 that RNA must be converted to cDNA before any downstream processing can occur. Performing RT before separation has no impact on the downstream processes and ensures that no information is lost to RNA degradation during separation. Both of these approaches are utilized in experiments presented later in this dissertation.

As a part of the Gel-Seq protocol, two devices were developed based on this underlying concept: a cassette based device and a glass slide based device. The cassette based device, fabricated on the macro-scale in standard gel cassettes, is designed to make this technology broadly accessible to researchers. The device can be fabricated in a standard biology lab without access to any specialized tools. A full protocol for the separation, recovery, and sequencing of a biological sample has been developed and is presented here. The performance of the device was validated by comparing the results to tube controls.

The glass slide based device shrinks the concept and to a device with features on the micron scale. By reducing the size of the device, it becomes possible to separate DNA and

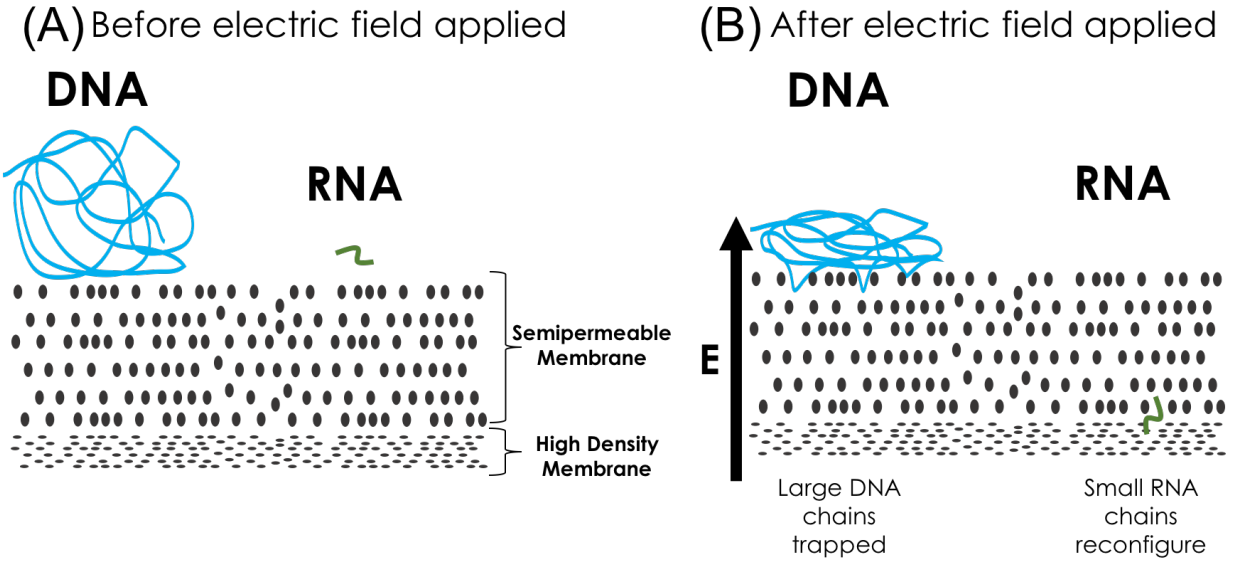


Figure 1.3: The underlying principle used to physically separate DNA and RNA. Under the presence of an electric field, small RNA molecules migrate through the membrane but large DNA molecules are trapped at the surface.

RNA from just a few hundred cells. The emphasis on this device was placed on developing the fabrication protocol. In this case, molds for casting the device must be fabricated using photolithography. Devices were tested to demonstrate that DNA and RNA could be separated at this reduced size scale. Future research will focus on rigorous validation of these devices in the same way as was done with the cassette based devices.

1.5 Dissertation Outline

This dissertation is written for the mechanical engineer with limited background in genetics and bioengineering. This introductory chapter has provided context for the basics in both cell biology and genetic sequencing. The next chapter covers the fundamental principles of electrophoresis, the core physics at the heart of the separation approach taken for DNA and RNA. This chapter introduces the possible material options that could be used for the electrophoretic separations and focuses on the material that was eventually selected: polyacrylamide. The third chapter focuses on the actual development of the Gel-Seq protocol. This chapter begins by showcasing the validation experiments that were performed during the design process. This is followed by a detailed examination of the two devices developed for use with the Gel-Seq protocol. The chapter concludes with a detailed look at the development of the biological protocols developed for the Gel-Seq protocol.

Chapter 4 reports on the optimization and validation of the Gel-Seq protocol. This

includes optimization of the membrane chemistry to physically separate DNA and RNA as well as the protocol for recovering DNA and RNA from the devices. The protocol is then evaluated by comparing the performance for a known standard to data generated with experimental controls performed in tube. The last chapter summarizes the dissertation and examines areas for future research.

Chapter 2

Fundamental Principles of Electrophoresis

Electrophoresis refers to the induced migration of charged particles in a liquid medium due to the presence of an electric field. No group has been more influenced by this phenomenon than molecular biologists. Once researchers discovered that electrophoresis performed in a hydrogel could be used to separate DNA fragments based on size, it became possible to investigate gene organization and expression. This lead to an intensive research effort in developing gel electrophoresis so that researchers could resolve minute size differences between DNA fragments ranging from 100 - 10,000 base pairs. Today, nearly every lab studying DNA or RNA contains standardized equipment to make these measurements. Indeed this technology has become so ubiquitous that the general public has been exposed to gel electrophoresis through popular television shows like C.S.I.

The research into electrophoresis for DNA and RNA has focused on the small fragments that are relevant for investigating genes. The work presented in this dissertation, on the other hand, is concerned with a new problem: separating small RNA or cDNA fragments ($< 10,000$ nucleotides/basepairs) from large fragments ($> 50 * 10^6$ basepairs).

This chapter focuses on introducing the core concepts underlying electrophoresis. First, the key governing equations are introduced for the electrophoresis of a particle. This analysis is then extended for electrophoresis of DNA and RNA. These two section present only an overview. For for more in depth treatment, refer Jean-Louis Viovy's excellent review upon which these sections were based: *Electrophoresis of DNA and other polyelectrolytes: Physical mechanisms* [23]. Following an introduction to the theoretical underpinnings of electrophoresis, an overview of material options for electrophoretic separations are introduced. The chapter concludes with a more in depth look at polyacrylamide, the material used to fabricate the devices presented in this dissertation.

2.1 Electrophoresis of a Particle

Before discussing the more complicated case of DNA gel electrophoresis, it is helpful to start by considering a simpler case: the electrophoresis of a spherical charged particle in an ionic solution. At its heart, electrophoresis is caused by the electrostatic interactions between the particle wall and the viscous forces in the fluid. Recall that for an electric field to be present in solution, there must be a sufficient number of ions present to carry a current. De-ionized water is a poor conductor, but sea water is an excellent conductor.

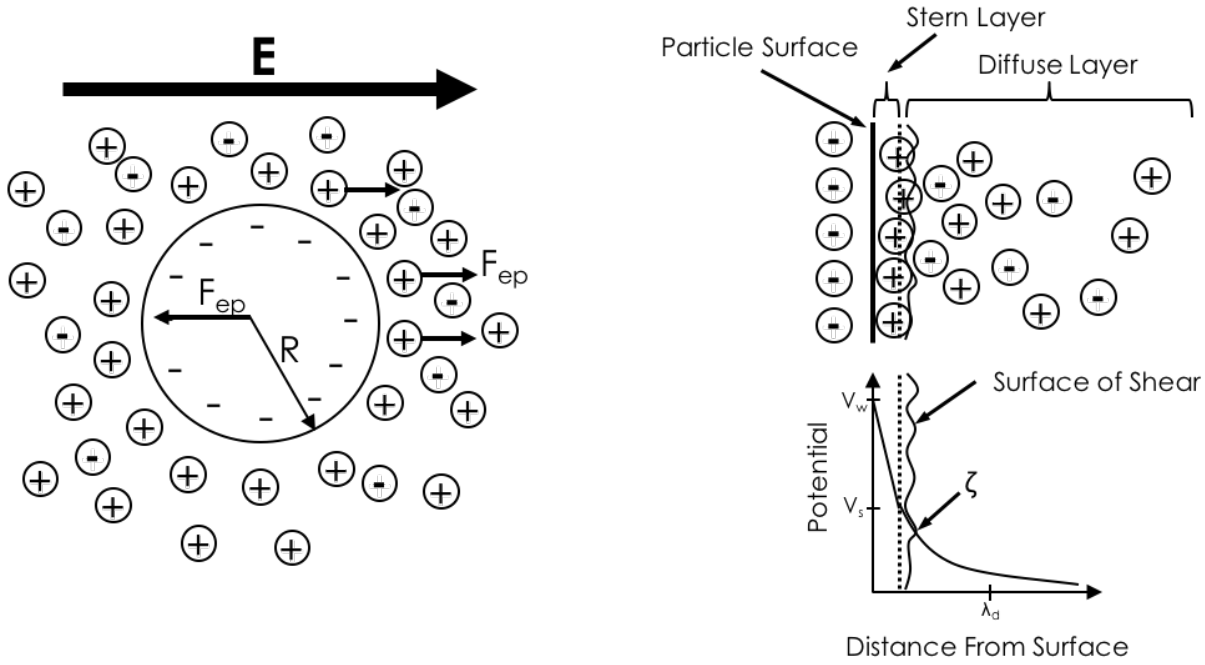


Figure 2.1: The distribution of ions for a charged particle in solution. The panel on the left shows some of the forces due to the application of an electric field. The panel on the right shows a close up of the particles surface. Images adapted from Viovy (left) and Probstein (right) [24, 23].

Consider the negatively charged particle shown in Figure 2.1. The application of an electric field E induces an electrophoretic force, F_ep on the particle. The ions in solution also experience an electrophoretic force, F_ep.

The negative charge of this particle perturbs the normal distribution of ions in solution and attracts a boundary layer of positive ions, shielding the charge of the particle at the wall (V_w). The equilibrium charge distribution that is reached is a balance between electrostatic and brownian forces. This layer of ions is known as the electric double layer and can be divided into two regions: immobile ions close to the surface (stern layer) and mobile ions further from the surface (diffuse layer). (See Figure 2.1.)

To understand the velocity U of this particle, it is useful to define the electrophoretic mobility, μ_{ep}:

$$\mu_{ep} = U/E \tag{2.1}$$

where E is the applied electric field. In order to model electrophoretic mobility, it is important to understand the distribution of ions in the diffuse layer. This distribution of ions can be modeled using the Poisson-Boltzman equation, for which the general solution is known as the Gouy-Chapman model. This model can be linearized to a simpler version known as the Debye-Hückel approximation. According to the Debye-Hückel theory, the electric potential due to the particle, $V_p(r)$, decreases exponentially from the maximum potential at surface of the particle V_s as:

$$V_p(r) = V_s e^{-\lambda_d r} \tag{2.2}$$

where λ_d is the Debye screening length, defined as:

$$\lambda_d = \sqrt{\frac{\epsilon RT}{2F^2 I_c}} \tag{2.3}$$

where ϵ is the electrical permittivity, R is the universal gas constant, T is temperature, F is Faraday's constant, and I_c is ionic strength of the solution. The Debye length is a characteristic length scale that indicates how far the electrostatic effects of the charged particle perturb the surrounding solution. It is useful indicator when trying to understand the electrophoretic behavior of a particle in solution.

Using the Debye-Hückel approximation, it is possible to consider two limiting cases for the electrophoresis of a spherical particle. First consider the case of a thick Debye layer, where $\lambda_d >> R$. In this case, the particle can be treated as a point charge and the electric force and viscous drag forces solved for separately. Equating these force and solving for the electrophoretic mobility gives the following relation:

$$\mu_{ep} = \frac{q}{6\pi\eta R} \tag{2.4}$$

where η is the viscosity of the fluid and q represents the net charge of the particle and the electric double layer. This result makes intuitive sense as a balance of the viscous and drag forces, however few particles are accurately modeled by these assumptions.

In the other extreme, consider the case of a thin Debye layer, where $\lambda_d << R$. In this case, it is impossible to independently solve for the forces and the Stokes equation $\eta\Delta V_p = -\rho E$ must be used to solve for the velocity where ρ is the solution of the Debye-Hückel model inside of the Debye layer and 0 outside. This analysis leads to the following result:

$$\mu_{ep} = \frac{\epsilon\zeta}{\eta} \tag{2.5}$$

where ζ represents the potential at edge of the stern layer and should not be confused with the wall or surface potentials of the particle. This analysis leads to the surprising result that the electrophoretic mobility for a smooth particle is independent of particle size and shape. Instead the driving factors are related to the viscosity of the fluid and the electrical interactions between the particle and ions in the fluid. With this understanding of the behavior of charged particles, now let us consider the more complicated case of DNA and RNA.

2.2 Electrophoresis of Nucleic Acids

DNA and RNA are long, negatively charged polymer chains and exhibit unique properties in the presence of an electric field. In order to understand the electrophoresis of these molecules, it is helpful to review some basic theory regarding polymer conformations. A polymer of length L can be treated as a combination of flexible segments, each with a persistence length l_p. Generally, including for DNA and RNA, this persistence length is longer than a single monomer unit.

In free solution, a polymer will not remain in an entropically unfavorable stretched state, instead the polymer will ball up into a more coiled state. In an uncharged polymer, the chain conformation can be described with a random walk analysis based on segments of length $l = 2 * l_p$. (This new variable l is known as the Kuhn length.) This analysis gives a mean end to end distance, R_N of the polymer as:

$$R_N = \sqrt{6} R_G = \sqrt{lL} \tag{2.6}$$

where R_G is the radius of gyration and describes the RMS distance of the various atoms from the center of mass of the polymer coil. When chains get longer, a simple random walk analysis is not sufficient to describe polymer conformations. In this case it is important to also consider the effects of steric hindrance where two monomer units cannot physically occupy the same space. This excluded volume theory is useful for DNA chains in excess of 10,000 basepairs, while the random walk analysis is useful for shorter chains [23].

In the case of DNA and RNA, which contain two negative charges per monomer, the brownian random walk analysis must take into consideration electrostatic forces. The polymer conformation thus becomes a function of the Debye length, persistence length, overall length, and monomer diameter. This analysis is further complicated by the fact that the Debye length itself is a function of the buffer conditions. While it is possible to tease out the impact of each of these factors, the end result is the same: DNA and RNA in free solution behave like a charged particle with a thin Debye layer. Recall from the previous section, the electrophoretic velocity of a charged particle with a thin Debye layer is independent of size (see Equation 2.5). Therefore DNA and RNA cannot be separated based on size using electrophoresis in free solution.

The solution to this problem is to perform electrophoretic separations of polyelectrolytes inside of a hydrogel matrix, a protocol known as gel electrophoresis. Hydrogels are polymer

matrices that are able to absorb and hold solvent. Their commercial uses range from soft contact lenses to diapers. When DNA and RNA undergo electrophoresis through these hydrogel matrices, it becomes possible to separate the molecules by size. The gel, unlike free solution, provides steric hindrance that changes the electrophoretic mobility of polyelectrolytes based on size.

The underlying physics that explains this phenomenon is a model known as reptation. At its most basic level, the reptation model treats DNA moving through a hydrogel as a snake moving through thick grass. In order to take advantage of this phenomenon, the hydrogel must have pores smaller than the radius of gyration of the polymer. This condition forces the polymer to uncoil and behave like a snake moving through the hydrogel. Short DNA molecules can easily reconfigure and move quickly through the polymer matrix. Long molecules, on the other hand, stretch out and come into contact with more gel. This increased contact between the DNA and gel makes it take longer for the DNA to reconfigure and move through the gel.

By changing the properties of the hydrogel, it is possible to achieve size separations at different length scales. There are two primary materials used in standard gel electrophoresis, agarose and polyacrylamide. Agarose is used to separate large DNA fragments. With low density agarose gels it is possible to separate DNA fragments up to 60,000 base pairs in length, though more commonly gels are used to separate fragments in the range of 500 - 30,000 basepairs [25]. Polyacrylamide hydrogels are useful for separating smaller DNA and RNA fragments. High density polyacrylamide gels can be used to separate fragments as small as 5 base pairs long, but gels are more commonly used to resolve size differences between 100 - 2000 basepairs in length [26].

Research into gel electrophoresis has focused on developing gels that are able to resolve small size differences. The ability to distinguish between an RNA transcript that is 1500 base pairs versus 1550 base pairs is quite useful for molecular biologists. The problem for this dissertation, however, is to separate chromosomal scale DNA from short RNA transcripts. Separating DNA and RNA of such different size scales has never been addressed by the electrophoresis community as it was not thought to be relevant.

The electrophoresis of DNA and RNA is also impacted by the buffer conditions used during separations. The buffer serves two primary functions in electrophoresis. First, the buffer maintains a constant pH that protect DNA and RNA from degrading. (When placed in acidic or basic solutions, DNA and RNA begin to denature and degrade.) Second, the buffer provides ions that carry current through the gel. Without the flow of this current, there would be no electrophoretic migration.

Careful control of the buffer conditions is required to ensure proper electrophoretic separations. If the concentration of the ions in the buffer is too high, the gel will carry large amounts of current and undergo substantial Joule heating. This heating can have detrimental effects on both they hydrogel and the sample. Hydrogels, like most polymers, have temperature sensitive properties. In the case of agarose, excessive Joule heating can actually cause the gel to melt during electrophoresis. Similarly, sufficient heating can also cause the DNA and RNA to denature. At the other end of the spectrum, should a buffer contain too

few ions, there will be no electrophoretic migration of the DNA and RNA [27].

Tris-borate-EDTA (TBE) and Tris-acetate-EDTA (TAE) emerged in the 1970s as the favored buffers for DNA and RNA electrophoresis [28]. One of the reasons Tris has become ubiquitous in electrophoresis buffers is that at the desired pH for DNA and RNA electrophoresis, it is in a partially uncharged state. (This is due to its base equilibrium being close to the ideal pH for separations.) With only a minority of the Tris ions charged, it is possible to use of high concentrations of buffering agents without the negative effects of increased current flow [29]. This makes these buffers ideally suited for preserving the integrity of DNA and RNA during long electrophoretic separation experiments.

The fundamental concept of DNA and RNA electrophoresis can be described simply as the migration of a charged molecule in the presence of an electric field. As this section has shown, however, digging even one layer deeper into the physics of these migrations requires an understanding of electrodynamics, hydrodynamics, and polymer physics. Having now developed this deeper understanding, the next section will explore in more detail the material options available for the electrophoretic separations required by the Gel-Seq protocol.

2.3 Material Options for Electrophoretic Separations

Recall from the introduction, the proposed device contains two electrophoretic membranes. The first membrane stops genomic DNA but allows RNA to pass through. The second membrane stops RNA transcripts, but allows buffer ions to pass through. As part of designing the devices described in this dissertation, a thorough review of the possible materials that could be used to fabricate these membranes was conducted. The results from this review are presented in Table 2.1.

Polyacrylamide, as mentioned in the previous section, is a common material used for gel electrophoresis. This material is straightforward to fabricate. Powder monomer and cross linker are mixed in a buffer solution such as TBE and then a free radical initiator is introduced to begin polymerization. By varying the concentrations of monomer and cross linker, it is possible to fabricate a range of gels with widely varying properties. While polyacrylamide gels are typically used to separate short DNA and RNA transcripts, it seemed possible the properties could be tuned to fabricate both membranes required in the device. The downside to the material is that the monomer, acrylamide, is a known neurotoxin and additional safety precautions must be taken until the material is polymerized.

Agarose gel, also mentioned in the previous section, is a standard material used for electrophoresis. Agarose gels are easy to cast: agarose powder is mixed with liquid, boiled until it is fully dissolved, poured into a mold, and allowed to cool to room temperature. (This protocol is nearly identical to another process readers will be familiar with: making Jello.) The major drawback to agarose is that it contains relatively large pores. While the material could likely be used to stop genomic DNA and allow RNA to pass through, it seemed unlikely agarose could be used to stop small RNA transcripts all together.

Table 2.1: Material options for electrophoretic separations in the Gel-Seq devices.

	Polyacrylamide	Agarose Gel	Cellulose Acetate	Nanoporous Silicon
Applications	RNA separation	DNA separation	molecular filtration	molecular filtration
Pore Size	20 nm – 200 nm*	200 nm – 500 nm	200 nm – 5 µm	10 nm – 50 nm
Fabrication	Free radical polymerization	Cast from boiling solution	Solvent casting	Etching
Pro	Widely tunable properties	Easy to work with	Chemically robust	Controllable features
Con	Neurotoxin	Large pores	Difficult to fabricate	Difficult to fabricate

*Polyacrylamide doesn't have true pores, the polymer chains dynamically sample space within the hydrogel matrix. This is an estimate of the effective pore size for comparison to the other materials.

Cellulose acetate is another material that has been used for the electrophoresis of DNA and RNA [30], however it is more commonly found in molecular filtration applications. Cellulose acetate membranes can be made over a wide range of porosities, however the fabrication is challenging and not well documented in the literature. (The fabrication processes seem to be closely guarded trade secrets of the manufacturers.) One major advantage of cellulose acetate is that it is chemically quite robust.

Nanoporous silicon is another a useful material for molecular sieving [31]. Unlike the other materials mentioned in this section, nanoporous silicon is a relative newcomer as well controlled fabrication protocols have only recently been developed. This technology provides the most control over pore size, but is also the most challenging to fabricate. Nanopores ranging from 10 - 50 nm can be chemically etched through silicon substrates. The resulting filters can be accurately characterized, but are fragile and expensive to fabricate. While a viable option, the fabrication difficulties make it the least appealing option.

Based on this review of material options, polyacrylamide was determined to offer the most promising path forward. Indeed, as will be described in Chapter 3, two polyacrylamide gel formulations were successfully developed: one to separate genomic DNA from RNA and one to stop small RNA molecules.

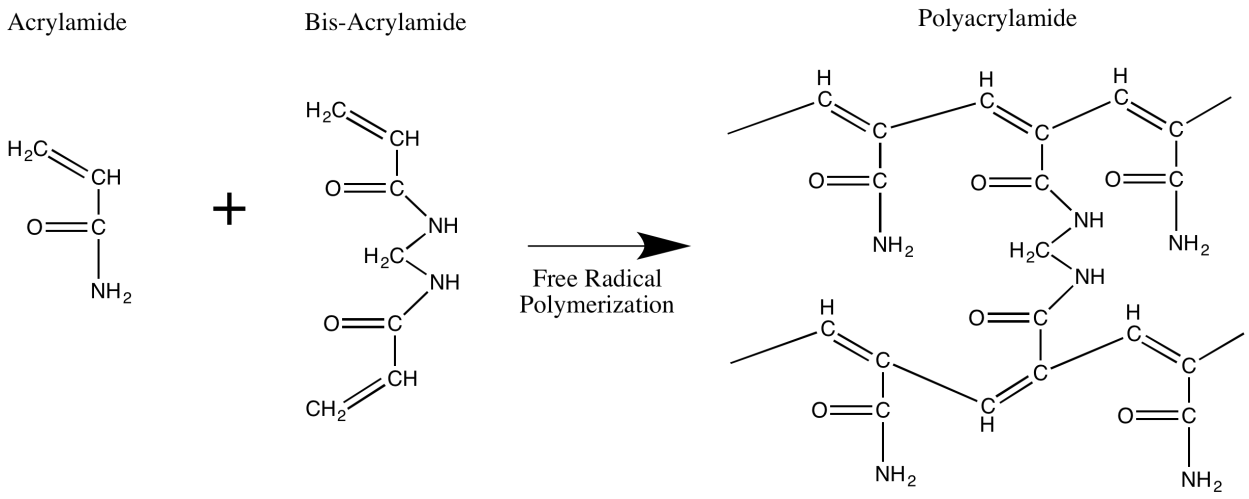


Figure 2.2: Molecular structures and the polymerization process for polyacrylamide.

2.4 Properties of Polyacrylamide

The devices described in this dissertation have been manufactured using polyacrylamide hydrogels. Polyacrylamide is an extremely versatile hydrogel that can modified to exhibit dramatically different properties. This section presents an overview of the important parameters that must be considered when fabricating devices using this material.

Polyacrylamide hydrogels are complex three dimensional matrices held together by both cross-linking and chain entanglement. Polyacrylamide is a linear homo-polymer composed of acrylamide monomers. These polymer chains can be cross-linked together using a range of different molecules, however the most common cross-linker used is bis-acrylamide. The chemical structures of the monomer units and final polymer are shown in Figure 2.2. Polyacrylamide hydrogel matrices have a polydispersed structure due to the fact that bis-acrylamide more readily polymerizes with itself than the acrylamide monomers. This results in regions with high concentrations of bis-acrylamide and regions of sparsely cross-linked acrylamide chains [32].

By varying the concentration of both monomer and cross-linker it is possible to fabricate a wide range of gels. Gel composition is defined using two terms %T and %C.The ratio between the total weight of acrylamide and bis-acrylamide cross-linker and the solvent in which it is dissolved is expressed as the weight/volume ratio %T. This parameter indicates the overall gel density. Gels below 3%T are too soft to handle easily, while gels above 50%T are difficult to fabricate. (Gels at 50%T approach the solubility limit of acrylamide and bis-acrylamide.) Polyacrylamide has been found to be useful for separating DNA and RNA fragments with gel densities ranging from 3.5 %T - 20 %T. Depending on the fragment sizes of interest, different gel densities are used to resolve size differences between fragments, see Table 2.2.

$$\%T = \frac{monomer\ mass(g) + crosslinker\ mass(g)}{solvent\ volume(mL)} \tag{2.7}$$

Table 2.2: Polyacrylamide gel densities for separating DNA and RNA fragments of different sizes [26].

Gel Density (%T)	Fragment Size (base pairs)
3.50%	1,000 bp - 2,000 bp
6.00%	75 bp - 2,000 bp
10.00%	30 bp - 1,000 bp
12.00%	40 bp - 200 bp
15.00%	25 bp - 150 bp
20.00%	5 bp - 100 bp

The weight ratio of cross-linker to total mass of monomer and cross linker combined is expressed as %C. The more cross-linker added to the gel, the more links created between the linear chains. One would reasonably expect that the more cross linker added, the harder it becomes for DNA and RNA to make its way through the gel. As expected DNA mobility decreases from 0.5%C to 5%C, however above 5%C DNA mobility actually begins to increase. This odd behavior is thought to be due to the fact that excess cross linking actually leads to a stiffening of the matrix and the creation of macrovoids that allow DNA to pass through the gel more quickly [33]. For this reason, gels used in nucleic acid separations are typically cross linked at either 3%C or 5%C depending on the application.

$$\%C = \frac{crosslinker\ mass(g)}{monomer\ mass(g) + crosslinker\ mass(g)} \tag{2.8}$$

Gel properties are affected not only by the monomer and cross linker concentrations, but also by the polymerization process. Polyacrylamide is polymerized through free radical polymerization. When added to monomer solutions, free radical initiators begin the chain reaction that leads to polymerization. The concentration of free radicals introduced into the gel can dramatically influence polymer chain length and the resulting gel properties. High concentrations of initiators lead to the formation of short polymer chains (low molecular weight), while low initiator concentrations lead to long chains (high molecular weight) [34]. The polymerization temperature also impacts reaction kinetics. Polymerization reactions conducted at high temperatures result in shorter chains than those conducted at low temperatures.

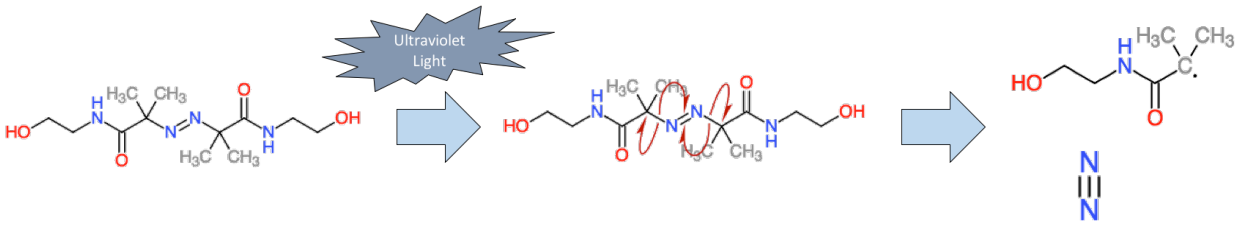


Figure 2.3: The mechanism by which free radicals are generated when azo initiators are exposed to UV light.

Free radicals initiators can be introduced into the system in a number of ways. The devices made in this dissertation were fabricated using two different polymerization approaches. The first approach, used for the cassette based devices, initiated polymerization with ammonium persulfate (APS). APS initiates polymerization at room temperature when the persulfate anion splits apart into two free radicals:

$$S_2O_8^{2-} + heat \rightarrow 2SO_4^{\bullet -} \tag{2.9}$$

In order to stabilize these free radicals, tetramethylethylenediamine (TEMED) is commonly added to the polymerization reaction. While not strictly necessary, TEMED acts as a catalyst and accelerates the polymerization reaction. While the reaction time depends on a range of factors, generally gels will fully polymerize between 30 min and 2 hours after the introduction of the initiator. Gels can be polymerized using small amounts of APS, only 1.25 mg/mL is required to polymerize a standard 6%T gel.

A second approach to initiating polymerization, used for the glass slide based devices, is to generate free radicals using ultraviolet (UV) radiation. Compounds known as azo initiators, such as 2,2'-Azobis2-methyl-N-[1,1-bis(hydroxymethyl)-2-hydroxyethl]propionamide (trade name VA-086, Wako Chemical), generate free radicals when exposed to UV light. The mechanism for the generation of these free radicals is shown in Figure 2.3. One advantage to azo initiated polymerization is the rapid pace at which gels polymerize. Using a high intensity UV source, $\sim 10\ mW/cm^2$, it is possible to polymerize gels in less than one minute. Another advantage to UV based initiation is that polymerization can be spatially controlled. Duncombe has developed a method for using photomasks to prevent certain regions from polymerizing, thereby creating features within polyacrylamide gels [35]. VA-086 is used to initiate polymerization at concentrations of $\tilde{1}0$ mg/mL.

This section has highlighted the relevant properties of polyacrylamide considered when designing the Gel-Seq devices. Nomeclature for describing gel density, %T and %C, was introduced. The free radical polymerization process was discussed and two mechanisms for initiating polymerization, APS and VA-086, were presented. This background will help the reader understand the experiments presented in the next chapter.

2.5 Chapter Summary

This chapter provided an overview of the fundamental principles in electrophoretic separations. Governing equations were developed for the simplest case of electrophoresis: a uniformly charged particle in a conducting medium. Key parameters, such as the Debye length and electrophoretic mobility, were introduced and explained. This analysis was used to set the stage for examining the electrophoresis of nucleic acids. The challenges of achieving size separations of DNA or RNA in free media were explained. This motivated the need for the development of gel electrophoresis, which provides the steric hindrance necessary to achieve the desired size separations. The different material options were discussed in reference to the functional requirements of the Gel-Seq protocol. The chapter concluded with a more detailed look at polyacrylamide, the material used to fabricate the Gel-Seq devices.

Chapter 3

Gel-Seq Development

This chapter reports on the successful development of an approach (Gel-Seq) to simultaneously sequence the genome and transcriptome from just a few hundred cells. The key innovation in this protocol was the development of devices that can physically separate DNA and RNA based on size. Two devices have been developed as a part of the Gel-Seq protocol, they are shown in Figure 3.1. While these devices have different form factors, which have different advantages, the fundamental operating principles are identical. The devices consist of two functional areas: a low density polyacrylamide membrane and a high density polyacrylamide membrane. In the presence of an electric field, DNA and RNA loaded into the device experience an electrophoretic force and begin to migrate. The low density region stops genomic DNA, but allows small RNA transcripts to pass through. These transcripts continue to migrate until they encounter the high density polyacrylamide, where they become trapped. Both membranes allow small buffer ions to pass through unimpeded, carrying current and setting up the conditions for electrophoresis.

In order to successfully develop the Gel-Seq protocol, the design process was started by identifying key objectives. The primary and intermediate objectives are reported in Table 3.1. The primary objectives identify the requirements to go from a sample to a matched genome and transcriptome library. The intermediate objectives identify important validation tasks that must be accomplished before the primary design objectives can be addressed.

As described in the previous chapter, polyacrylamide was selected as the best candidate for fabricating both the high density and low density membranes based on a thorough review of the available materials. The first section of this chapter presents results from experiments validating this material choice, addressing the intermediate design objectives. After validating polyacrylamide as a viable material option, work on the primary design objectives began. Two device form factors were developed: a cassette based device and a glass slide based device. The next sections of this chapter document the design and fabrication of these devices. These sections also highlight the strengths and weaknesses of each design. In addition to developing the devices, it was also necessary to develop a new biological protocol compatible with the physical separation achieved using the devices. The last section of this chapter covers the development of a biological protocol for use with the Gel-Seq protocol.

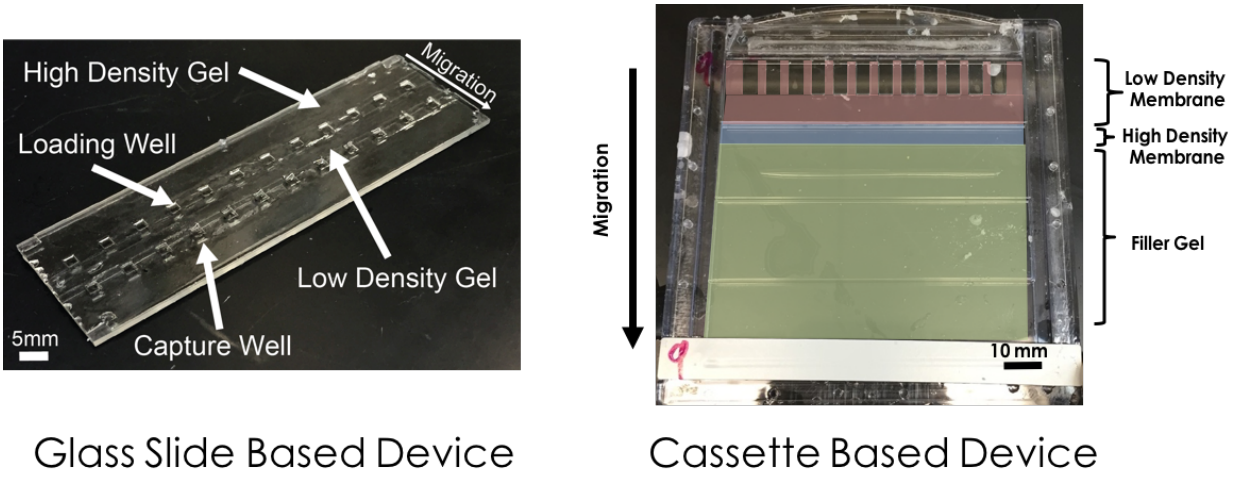


Figure 3.1: Two devices developed for separating DNA and RNA. The cassette based device is shown on the left, the glass slide based device is shown on the right.

Table 3.1: The primary and intermediate design objectives identified during the design process.

Primary Design Objectives
Develop a device to physically separate genomic DNA and RNA/cDNA
Demonstrate genomic DNA and RNA/cDNA can be recovered from the device
Adapt biological protocols to generate sequencing libraries using samples harvested from device
Minimize required amount of starting sample

Intermediate Design Objectives
Fabricate a membrane that stops genomic DNA, but permits RNA/cDNA and buffer ions
Fabricate a membrane that can be used to stop RNA/cDNA but permit buffer ions

3.1 Polyacrylamide Validation

The device design calls for the fabrication of two membranes, one high density and one low density. While the literature review suggested polyacrylamide was a viable candidate, there was no prior research demonstrating the material could be used in this way. Therefore a series of experiments were conducted in order to determine whether polyacrylamide membranes could be manufactured with the desired properties. Five proof of concept tasks were identified:

1. Move RNA/cDNA through low density gel
2. Stop genomic DNA from entering low density gel
3. Stop RNA/cDNA with a high density gel
4. Move RNA/cDNA from a loading well to a defined capture location
5. Understand electric fields in polyacrylamide gels

Task 1: Move RNA through low density gel

The first task was to demonstrate that RNA/cDNA could be moved through a low density gel. While RNA is commonly separated using low density gels, this task was important to demonstrate the author's competency in gel fabrication as well as validate the experimental setup. For this experiment, gels were fabricated using a protocol adapted from Duncombe [35]. Gels were polymerized using the UV initiator VA-086. Using UV polymerization made it possible to quickly fabricate and tune gel properties. Additionally features could be defined using inexpensive photomasks, making it possible to test multiple geometries.

The fabrication protocol is shown in Figure 3.2. Devices were fabricated on the substrate GelBond (Lonza), a functionalized polyester film that bonds to polyacrylamide. A cavity was created by placing a glass slide on top of 200 μm square latex spacers. The glass slide was treated with Gel Slick (Lonza), a silane solution that prevents polyacrylamide from sticking to the glass. The cavity was then filled with liquid polyacrylamide monomer solution containing acrylamide, bis-acrylamide, and the photo-initiator VA-086. (This gel contained 6%T,3%C with 1% w/V VA-086). A photomask containing three rectangles was placed on top of the glass slide. The entire assembly was then exposed to UV light, as shown in panel B, which induces polymerization in the regions not protected by the photomask.

The resulting devices are shown in Figure 3.3. Panel A shows two devices that have been placed in a 3D printed test fixture. The transparent devices are placed on a glass plate. Electrode wicks (Crescent Chemical) are placed at the top and bottom of the device. These wicks, essentially sponges, are soaked in the electrophoresis buffer TBE and act as reservoirs supplying buffer ions to conduct current during electrophoresis. Graphite electrodes are placed on top of the electrode wicks and are connected to a power supply (not shown).

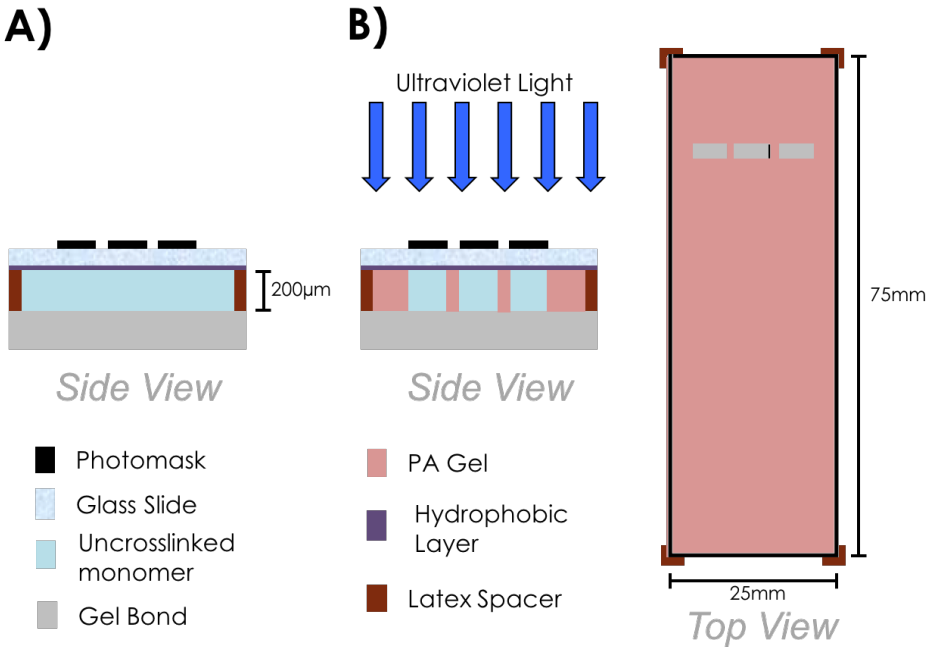


Figure 3.2: Initial fabrication protocol (generation 1) for making polyacrylamide gels.

.

The left image of Panel B shows a closeup of the device, the three loading wells have been highlighted with red boxes for easy visibility. Each loading well is 5 mm x 2 mm x 500 μm, with a total volume of 5 μL.

The goal in this task was to demonstrate the successful migration of short RNA/cDNA fragments in a polyacrylamide gel. In order demonstrate this behavior, RNA and DNA fragments known as ladders were used instead of a biological sample. Ladders are synthetically manufactured sequences with several distinct fragment sizes. They are used by biologists in standard gel electrophoresis as a control when estimating the size of unknown fragments. In this case, the ladders allow for the straightforward quantification of the migration behavior of different size fragments.

For this experiment, two different ladders were selected. The Low DNA Mass Ladder from Invitrogen was chosen to represent cDNA. This ladder contains six DNA fragment sizes: 100, 200, 400, 800, 1200, and 2000 base pairs. To represent RNA transcripts, the RNA 6000 ladder from Ambion was selected. This ladder contains six RNA fragments in the following sizes: 200, 500, 1000, 2000, 4000, and 6000 nucleotides. During testing, each lane of the device was loaded with a different ladder or combination of the two as shown in Figure 3.3. In order to visualize the progress during loading and migration, the samples were mixed with blue loading dye (Thermo Scientific 6X DNA Gel Loading Dye). 2 μL of the sample / dye combination was loaded into each well.

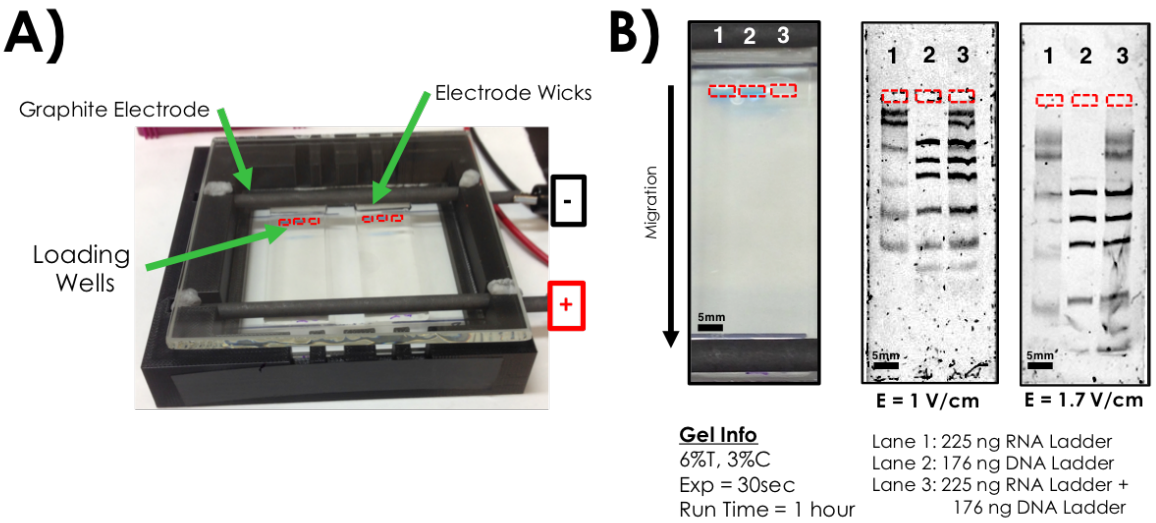


Figure 3.3: Devices demonstrating the migration of RNA through polyacrylamide gel. Panel A shows two devices placed in a 3-D printed test fixture. The image on the right shows a close up photograph of the device (left) and fluorescent images of RNA migration at two different electric field strengths (center and right).

After loading, an electric field was applied to the device to induce migration. Both the strength of the field and time the field is applied have an impact on the migration behavior. Multiple devices were tested and the results of two representative conditions, 1 V/cm and 1.5 V/cm for 1 hour, are shown in Figure 3.3. To visualize the location of the DNA and RNA, the nucleic acids are fluorescently stained and imaged using a UV camera. Unless otherwise noted, the devices in this dissertation were stained in 30 mL of 1X SYBR Gold Nucleic Acid Gel Stain (Invitrogen) for 5 minutes and imaged using a BioRad Gel Doc. The resulting image shows crisp black bands when DNA or RNA is present.

The images in Panel B of Figure 3.3 show that a 6%T, 3%C gel, exposed for 30 seconds to 11.1 mW/cm^2 ultraviolet radiation to initiate polymerization, is well suited for allowing DNA and RNA to move through the gel. Examining the center and right image, it is possible to see the six black bands associated with the different fragment sizes for each ladder. The larger DNA and RNA molecules migrate slowly and remain close to the loading wells. The shorter molecules move through much of the gel. As would be expected, when the electric field is increased (right image) the DNA and RNA ladders migrate further than the lower electric field (center image). This result demonstrated competency in gel fabrication as well as validated the experimental setup.

Task 2: Stop genomic DNA from entering low density gel

The ability to stop DNA from entering a gel is something that runs counter to research in polyacrylamide gel electrophoresis. Developing gel chemistry that would stop genomic DNA was an important step in validating the selection of polyacrylamide for the high density membrane. Using the fabrication approach described in the previous section, an 8%T, 3.3%C gel was polymerized with 30 seconds of UV at 11.1 mW/cm^2. A synthetic analog does not exist for genomic DNA, therefore this experiment required biological samples. Rather than use entire cells, which contain both DNA and RNA, nuclei were selected as they contain primarily genomic DNA. The device was loaded with 2 μL of a cells suspension containing approximately 2000 nuclei stained with SYBR gold. The device was then imaged to validate loading. The loaded well is shown on the left panel of Figure 3.4. The nuclei appear as punctate black dots.

In order to make the DNA within the nuclei accessible for electrophoresis, the nuclei were lysed by adding 1.5 μL alkaline lysis solution (400 mM KOH, 100 mM DTT, 10 mM EDTA) to the loading well and incubating at room temperature for 3 minutes. The resulting solution was neutralized by adding 1.5 μL of neutralizing solution (250 mM HCl, 666 mM Tris HCl). To confirm the gel would still permit small fragments to enter, an adjacent well was loaded with a DNA ladder. The gel was run for 1 hour with an electric field of 0.7 V/cm and then imaged. The results are shown in the right panel of Figure 3.4. The DNA ladder, as in the previous experiment, enters the gel and begins to migrate. The genomic DNA released from the nuclei, however, has piled up as a single crisp black band at the edge of the loading well. This result demonstrates that it is possible to fabricate a polyacrylamide that allows small fragments to enter, but stops DNA on the genomic size scale.

Task 3: Stop RNA/cDNA with a high density gel

In order to facilitate the recovery of RNA and cDNA, it is necessary to stop these molecules in a defined location. If it were not possible to stop the RNA, it would continue to migrate until it came into contact with the electrode and was destroyed. Similarly if the RNA were to become spread out over a wide section of gel due to the different electrophoretic mobilities of the various fragments, it would be very challenging to recover for sequencing. This experiment explored whether a high density polyacrylamide could be used to stop RNA. Making a high density polyacrylamide gel required the development of a new fabrication protocol for two reasons. First, residual stress in the high density gel caused the polymer substrate used in the first fabrication protocol to curl. Second, a photomask could not be used to define features in the gel. When the high density gel was polymerized, the regions beneath the photomask also polymerized due to the high monomer concentration.

The new fabrication protocol, Generation 2, is shown in Figure 3.5. In order to address the issue of residual stresses warping the device, the polymer substrate has been replaced with a glass slide. In order to promote the adhesion of the polyacrylamide to the glass slide, the surface was functionalized with 3-(Trimethoxysilyl)propyl methacrylate. This created

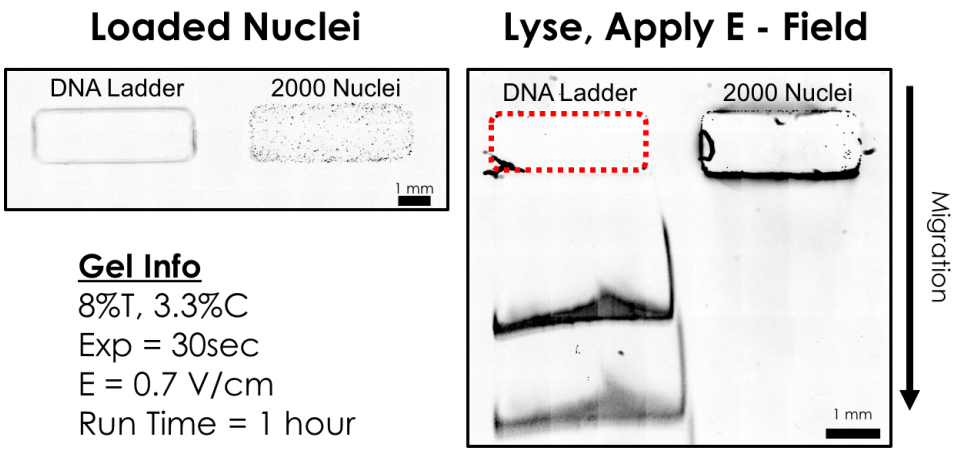


Figure 3.4: A fluorescent image showing that genomic DNA can be prevented from entering a low density polyacrylamide gel. The well on the left was loaded with a DNA ladder containing small DNA fragments, while the well on the left was loaded with the genomic DNA from nuclei. After electrophoresis the DNA ladder has entered the gel, but the genomic DNA has been stopped.

a strong chemical bond between the polymer and the substrate. To address the issue of unwanted polymerization, latex spacers were used to physically exclude the monomer solution from the loading wells. Essentially the latex spacers acted as a mold, defining regions where the liquid monomer could not flow. These two modifications made it possible to fabricate high density polyacrylamide gels.

To validate that the RNA and cDNA could be stopped with a high density gel, a 40$T, 5%C gel was fabricated. The device was tested using the same RNA and DNA ladders from Task 1. The gel was subjected to an electric field of 1 V/cm for 30 min, stained with a 1x SYBR gold solution, and imaged. The results are shown in Figure 3.6. A clear black band is evident at the edge of the loading well. This indicates that the RNA/DNA ladders experienced an electrophoretic force that pushed them out of solution onto the edge of the gel, but that they were unable to enter the gel. Notice in the figure that the three loading wells are neither uniformly sized or oriented, this is due to variations in cutting the latex spacers by hand as well as slight movement of the spacers during fabrication. This result shows that is possible to create a high density polyacrylamide gel that stops RNA and cDNA, but still permits ions to flow during electrophoresis.

Task 4: Move RNA/cDNA from a loading well to a defined capture location

The first three tasks were focused on validating polyacrylamide as a viable material. Task 4 examines whether membranes of two different densities can be fabricated in the same

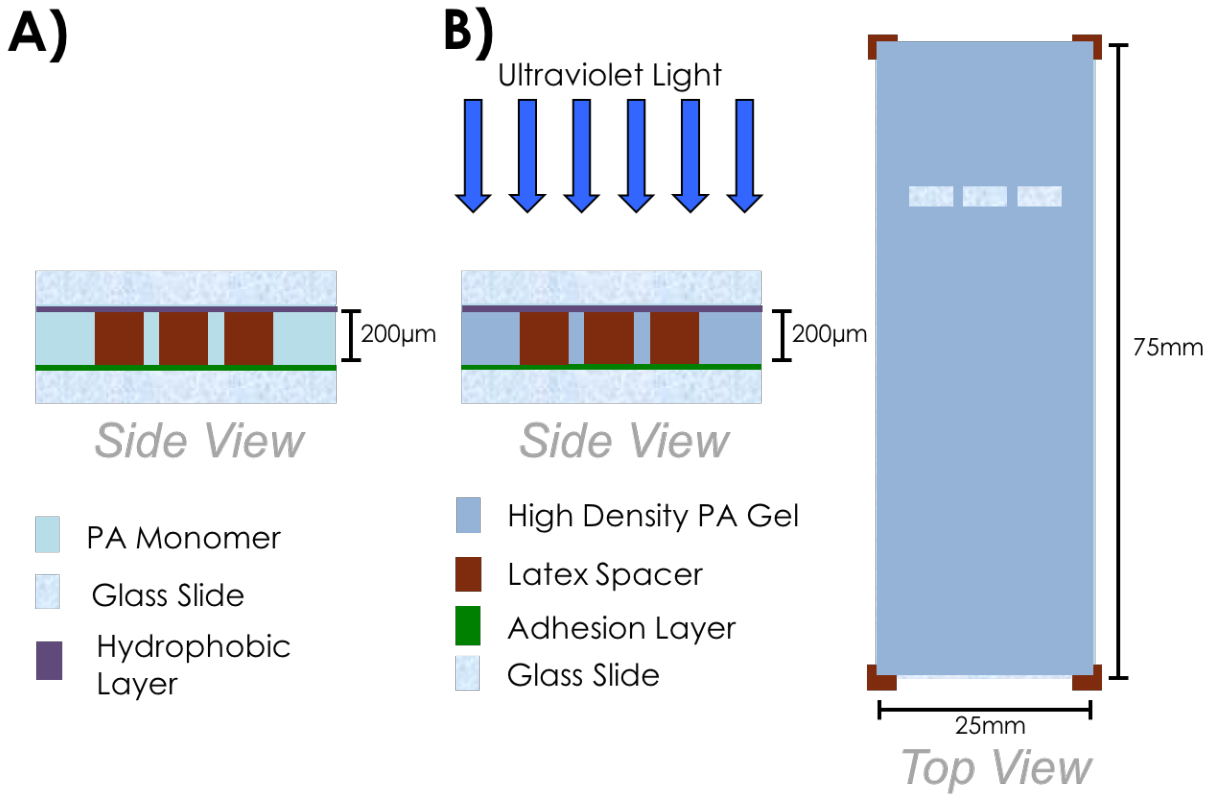


Figure 3.5: A revised fabrication protocol (generation 2) for making high density polyacrylamide gels.

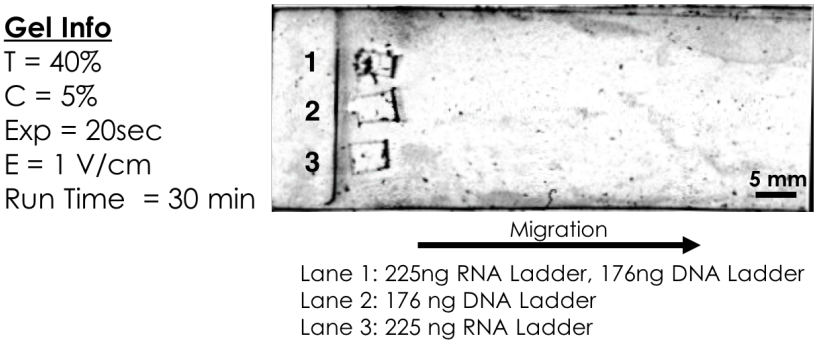


Figure 3.6: A fluorescent image showing that a high density polyacrylamide gel can be used to stop small fragments of DNA and RNA.

device. The objective in this experiment was to demonstrate that RNA could be loaded into a well, moved through a low density gel, and trapped at the other side by a high density gel. The major challenge addressed here was demonstrating the fabrication of devices with two different membranes.

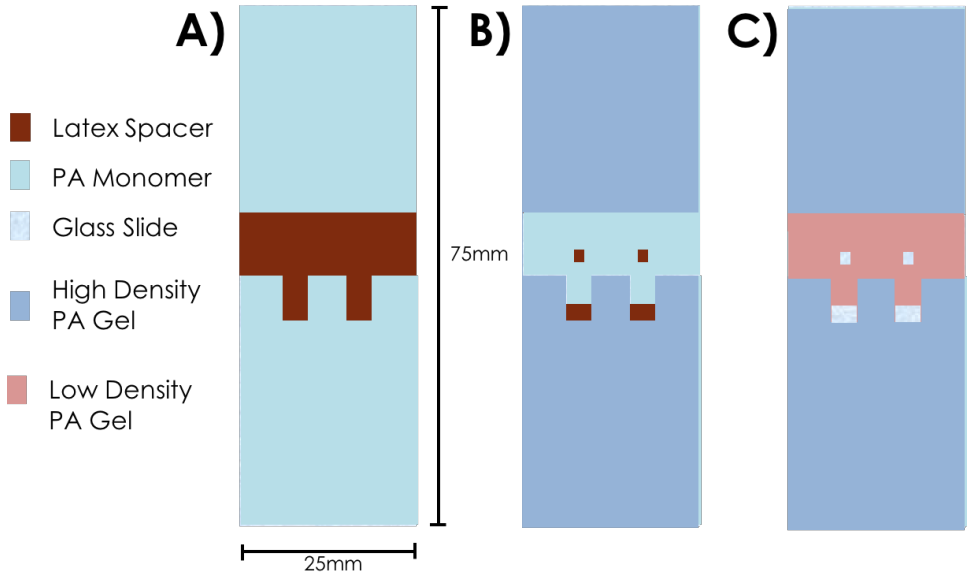


Figure 3.7: An overview of the fabrication protocol used to make devices with regions of both high and low density polyacrylamide. Panel A shows how a large latex spacer was used to define the high density region. Panel B shows how four small spacers were used to define the low density region. Panel C shows the regions of high and low density after polymerization.

A new fabrication protocol was developed to integrate the fabrication of a low density gel and high density gel on the same device. An overview of the protocol, generation 3, is shown in Figure 3.7. The protocol uses the same basic approach as the previous generations. A latex spacer is sandwiched between two glass slides, one designed to bond to polyacrylamide, the other designed not to bond to polyacrylamide. In the first step, shown in panel A, a high density liquid polyacrylamide monomer is pipetted between the glass slides. In this case, the latex spacer protects the region where the low density monomer will eventually be fabricated. The assembly is exposed to UV light, curing the high density monomer. The slides are then separated and the latex spacer removed. New latex spacers are then added to define a loading and capture well. A low density polyacrylamide liquid monomer is then pipetted between the glass slides to fill the remaining volume. The assembly is again exposed to UV, polymerizing the low density polyacrylamide. Note that the additional exposure of UV to the high density membrane does not seem to have a substantial impact on the polymers properties. Once polymerization of the low density membrane is complete, the cover slide and latex spacers are removed, leaving behind the completed device shown in Panel C.

In order to simplify experimental conditions, RNA and DNA ladders were again used to represent the RNA or cDNA from a biological sample. The ladders were loaded into the device, an electric field was applied for 30 min, the gel was stained, and the results imaged. Multiple gel chemistries were tested, the results in Figure 3.8 are from a 40%T, 5%C high density region and a 5%T, 3.3%C low density region. False color has been added to help identify the regions of high density gel. The results for 1 V/cm (left panel) show that while the ladders enter the gel, they do not fully pass through to be collected on the other side. Increasing the electric field to 1.6 V/cm, however, results in most of the ladder moving through the low density region of the gel. This result demonstrated that it is possible to fabricate adjacent regions of low and high density polyacrylamide and move small fragments of DNA and RNA from a loading well to a specific capture location.

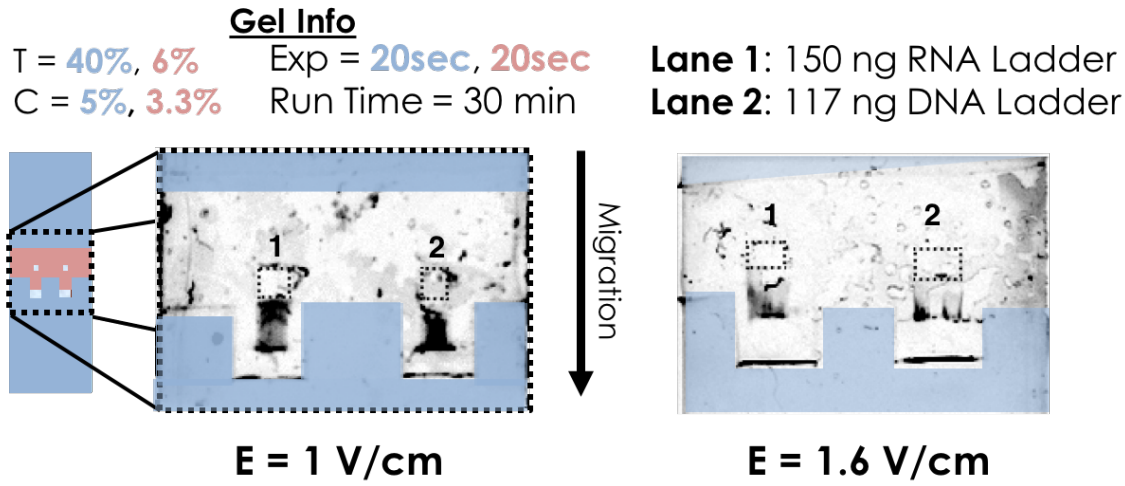


Figure 3.8: A device showing that RNA/cDNA can be moved from a loading well through a low density gel to a defined capture location. The device shown has been fabricated with both high and low density regions. While an electric field of 1 V/cm is not sufficient to drive the DNA and RNA ladders through the low density region, increasing the electric field results in most of the later moving through the gel.

Task 5: Electric Fields in Polyacrylamide

As was previously introduced in Chapter 2, the electrophoretic mobility of DNA and RNA is proportional to the applied electric field. In devices containing only a single gel density, there is a uniform electric field throughout the device. In devices with a high and low density membranes, however, it quickly became apparent that it was incorrect to assume there was a uniform electric field throughout the two membranes.

After observing slower than expected migration, a voltmeter was used to measure the potential difference across regions of both high and low density. This investigation revealed

that high density gel membranes have substantially higher resistivity than the low density gels. Intuitively this result makes sense: for a given volume high density gels contain more non-conductive polymer and, therefore, less conductive buffer.

Table 3.2: The inputs used for a model of the electric field in devices with two different densities.

Model Parameter	Value
High Density Gel Resistivity, ρ_H	9.8 Ωm
Low Density Gel Resistivity, ρ_L	2.3 Ωm
Electrode Wick Resistivity, ρ_W	300 Ωm
Applied Potential	200 V

In order to examine the electric fields within devices containing two different density membranes, a Comsol model was developed. Computational models are only as good as their inputs, therefore it was necessary to accurately measure resistivity values for the high and low density membranes. Two devices were made, one containing only 40%T,5%C polyacrylamide and one containing only 6%T,3%C polyacrylamide. Each device, using graphite electrodes and electrode wicks, was connected to a source measure unit (SMU) and a constant current of 1 mA was applied. The SMU was used to measure the potential different across the device. After accounting for the resistance of the electrode wicks (2 kΩ), the resistivity for each gel density was calculated as follows:

$$\rho = R * A/l \tag{3.1}$$

where R was the measured resistance, A was the cross sectional area of the device (12.5 mm^2), and l was the length of gel between the electrodes (65 mm). Using this method, the model inputs were defined as shown in Table 3.2. Having defined the model inputs, a physical model of the devices was built using Solidworks. This CAD model was imported to Comsol and the electric currents module was used to predict the electric field in the gel.

The results from the Comsol simulation are shown in Figure 3.9. The figure shows heat maps of the predicted electric potential (V/cm). The CAD models without results are also shown below the heat maps to reveal the locations of high and low density gel. High density polyacrylamide is shown in blue, the low density polyacrylamide in pink, and the electrode wicks in gray. The device used in this modeling is a slightly updated version of the two density gel discussed in the previous section.

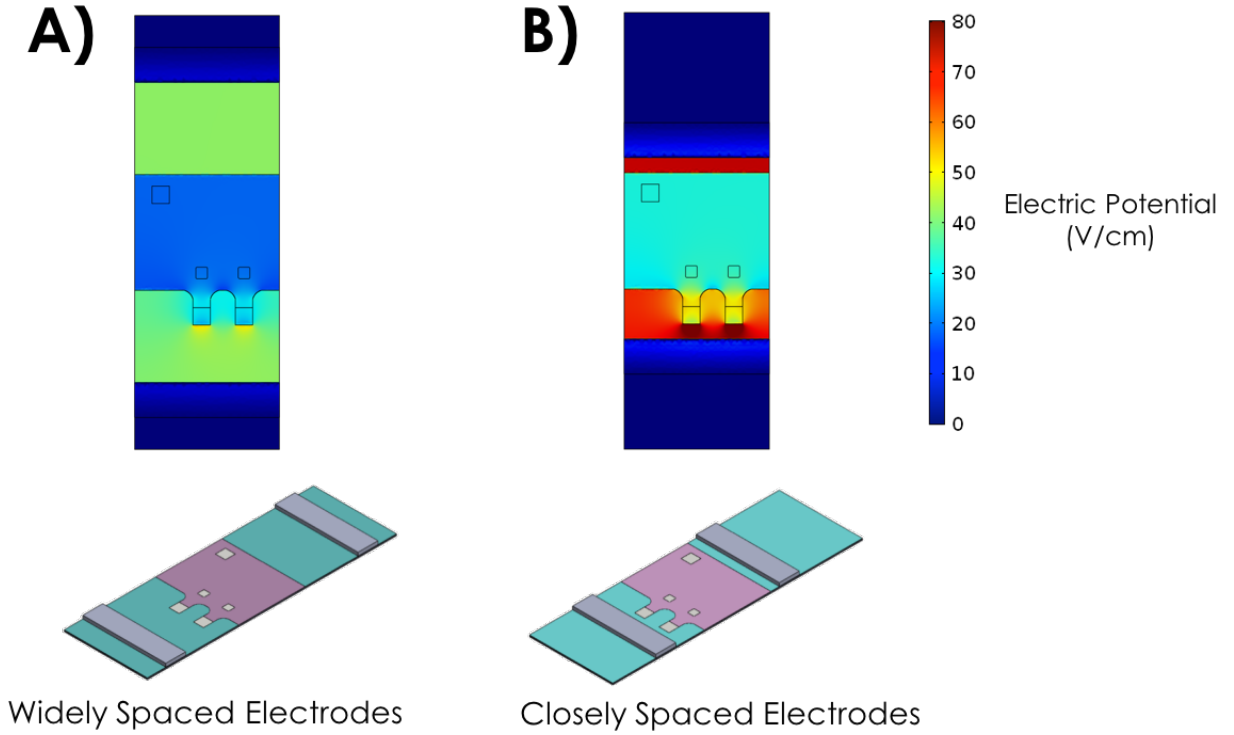


Figure 3.9: Two models of the electric field in devices with both high and low density regions of polyacrylamide.

Panel A of Figure 3.9 shows the simulation results for a condition where the electrode wicks are placed at the edges of the device. In this case, the electric potential in the low density gel, where the electrophoretic migration occurs, is close to zero. Instead most of the potential drop occurs over the high density sections of the gel, where no migration occurs. This result indicated that the final design should minimize the amount of high density gel in the conduction path. This would maximize the potential difference across the portion of the device where the separation of DNA and RNA was to occur. Panel B shows the impact of positioning the electrodes much closer to the low density gel. (This can be seen by examining the location of the grey electrode wicks on the CAD model below the heat map.) In this case, a much higher electric field is created in the region of interest.

The results of this modeling effort were experimentally confirmed. Changing the position of the electrode wicks allowed for higher control over the electric field in the device. This added control made it possible to drive the RNA ladder completely through the low density membrane. The differences between high and low density fields are shown in Figure 3.10. This understanding of the impact of large sections of high density gel on the device behavior was integral to developing the final designs discussed in the rest of this chapter.

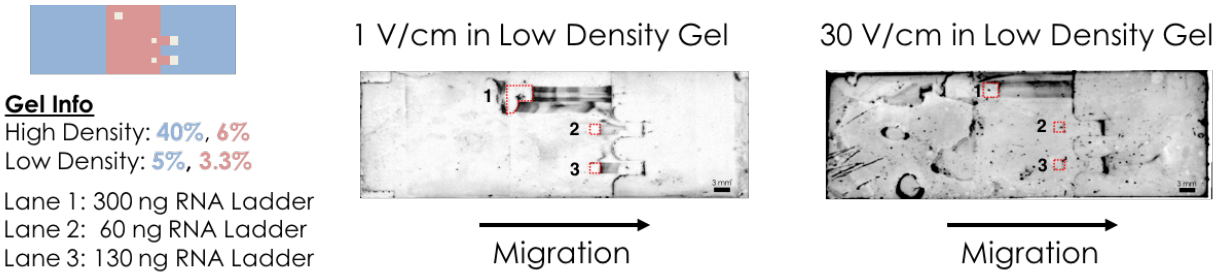


Figure 3.10: Repositioning the electrodes based on the results from the Comsol model gave much higher control over the electric field within the low density gel. This made it possible to identify the correct conditions to drive RNA ladder through the separation gel.

Polyacrylamide Validation Summary

The experiments described in this section confirmed the selection of polyacrylamide as a viable material for both the high and low density gels. By tuning the gel properties, it is possible to create membranes at opposite ends of the spectrum. For the low density membrane, polyacrylamide gels on the order of 6%T, 3.3%C stop genomic DNA but allow RNA/cDNA and ions to pass through. For the high density membrane, 40%T, 5%C gels stop RNA/cDNA but allow the flow of ions. Furthermore these experiments showed that these membranes could be fabricated next to each other on the same device. Several different fabrication protocols were developed that utilized the photo initiator VA-086 to rapidly polymerize the gel. Low density membranes could have features defined using a photomask, but high density membranes required a mold to prevent unwanted polymerization. These initial findings confirmed that polyacrylamide membranes could be used to achieve the required design objectives. By developing a computational model to predict the electric fields within the device, insight was gained into the importance of minimizing the amount of high density gel in the conduction path. Based on these experiments, two classes of devices were designed and fabricated: a glass slide based device and a cassette based device.

3.2 Glass Slide Based Device Design and Fabrication

One of the objectives in this project is to develop a protocol that can start from hundreds of cells. DNA and RNA are not lost during the process of gel electrophoresis, however there is the potential for loss when the sample is recovered from the gel. The recovery and testing protocols will be discussed in the next chapter, but essentially the objective is to minimize the amount of gel in which the sample is embedded. This is one of the reasons why a high density membrane was required to stop the RNA. Without this high density membrane, the RNA would be too spread out in the gel to be effectively recovered.

To satisfy this objective, a device was developed around the form factor of a glass microscope slide. This device, shown in Figure 3.11, consists of high and low density regions

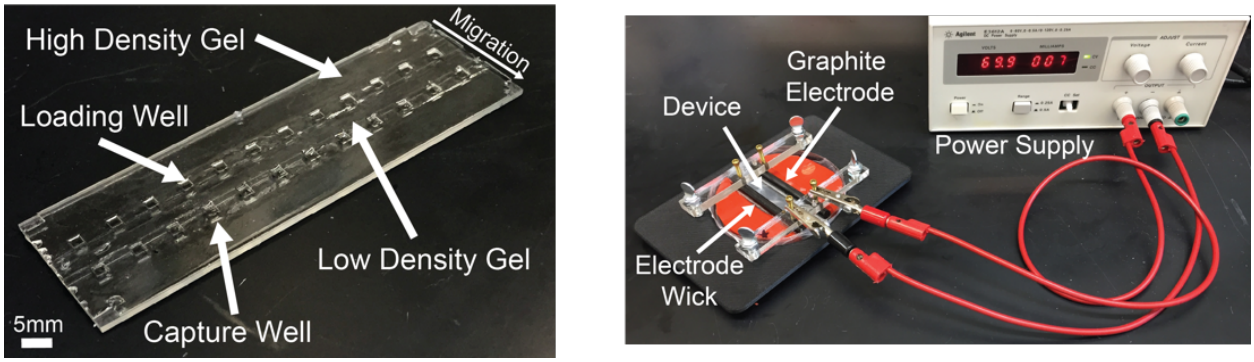


Figure 3.11: The glass slide based device (left) and the test setup used to apply an electric field to the device (right).

of thin polyacrylamide gel. In this case, the migration direction is perpendicular to that of the devices developed in the previous section, allowing for more samples to be tested in a single run. Each device contains 12 pairs of loading and capture wells. Biological samples are loaded into the wells (2mm x 2mm x 250 μm, $1\mu L$ in volume). Using the test fixture, also shown in Figure 3.11, an electric field is applied to the device inducing migration in the DNA and RNA. The DNA becomes trapped in the loading well, while the RNA is able to move through the low density gel and becomes trapped on the opposing wall of the capture well. Electrode wicks and graphite electrodes are used just as in the previous section to apply the electric field to the device, however in this case they are aligned along the long edge of the device.

While it is possible to fabricate and test multiple gel chemistries, the results presented in the next chapter are for a high density region of 40%T, 5%C and a low density region of 8%T, 3%C. The fabrication protocol for this device improves upon the methods presented in the previous section, an overview is shown in Figure 3.13. Fabrication is accomplished through a combination of photolithography and molding.

First a 250 μm thick SU-8 mold is created on a glass wafer. The SU-8 topology defines the negative space of the high density polyacrylamide region. Using SU-8 as the mold material has several advantages over latex spacers. First, the mold is permanently bonded to the glass wafer, eliminating the problem of the mold changing due to the flow of liquid monomer. Second, using photolithography allows the precise definition of small mold features. The latex spacers were cut with scissors and razor blades, making it a challenge to fabricate features with both accuracy and precision.

SU-8 molds were fabricated using the following procedure. A 100 mm glass wafer was cleaned for 5 min with oxygen plasma at 200 mW. The wafer was then placed inside a spin coater and SU-8 2100 (MicroChem) was poured into the center of the wafer. The spin coater was run at two speeds. The first speed was 500 rpm for 30 seconds with an acceleration of 100 rpm/second. This step helped to evenly distribute the SU-8 over the surface of the wafer. The second step, 1000 rpm for 30 seconds, acceleration of 300 rpm/second, was used

to thin the coating to the desired 250 μm. After spin coating, the edges of the wafer were cleaned using SU-8 developer. The wafer was then soft baked at 110°C for 60 min to remove the solvent from the SU-8. Wafers were then placed in a photolithography tool, covered with a photomask (See Figure 3.12, and exposed to 540 mJ/cm^2 of UV energy (equivalent to 41.5 seconds on the EVG-620). The wafer was then subjected to a post exposure bake of 16 minutes at 110°C. Finally the un-polymerized SU-8 was washed away by submerging the wafer in SU-8 developer (Micro Chem) and sonicating the solution for 4 - 5 minutes. After fabrication the mold was treated with Glass Free (National Diagnostics EC-621) to improve separation of the mold from polyacrylamide. The SU-8 molds were submerged in the Glass Free solution for 5 minutes. The molds were then removed and washed first using toluene and then methanol. Finally the molds were dried using compressed air.

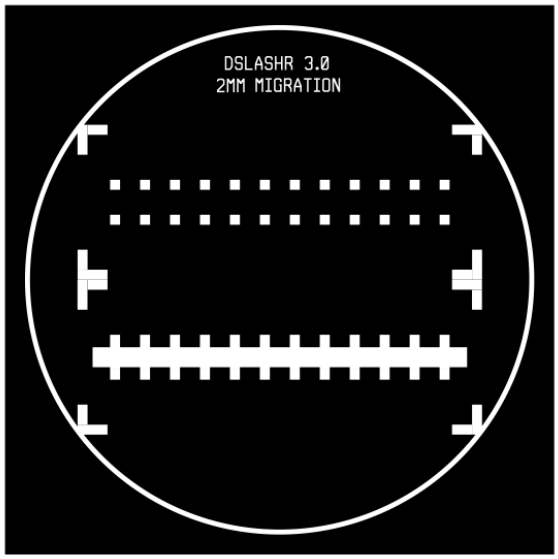


Figure 3.12: The photomask used to define the SU-8 features for the glass slide based devices.

The glass slide that is to be the substrate for the device is treated with 3-(Trimethoxysilyl) propyl methacrylate. This silane treatment is necessary for the creation of a covalent bond between the polyacrylamide and glass slide. In order to create a robust adhesion layer, slides were first cleaned using soapy water and dried on a hotplate. Slides were then immersed for 3 minutes in a solution containing 1 mL of 3-(trimethoxysilyl)propyl methacrylate, 200 mL of ethanol, and 6 mL of dilute acetic acid (1:10 glacial acetic acid:water). Slides were then rinsed using ethanol and dried with compressed air.

The treated glass slide is positioned 300 μm above the SU-8 mold using latex spacers at the corners. The mold cavity is filled with a polyacrylamide monomer solution (40%T, 5%C) containing the photo-initiator VA-086 (1% w/V). The entire assembly is exposed to UV light (13 mW/cm^2) for 45 seconds, resulting in rapid polymerization of the high density polyacrylamide region. Once polymerized, the glass slide and high density PAG layer are

separated from the SU-8 mold. This process creates a 300 μm layer of high density polyacrylamide in most regions of the device, and a 50 μm layer of high density polyacrylamide in regions where low density polyacrylamide will be added. The thin region of high density polyacrylamide is beneficial for two reasons. First it minimizes air bubble formation during device fabrication. Second it facilitates the adhesion of the low density polyacrylamide to the device. (The low density polyacrylamide has better adhesion to high density polyacrylamide than to glass.)

The molding process is then repeated using a new 250 μm SU-8 mold to define the loading and capture wells. For the second molding process, the high density polyacrylamide serves as the spacer to suspend the glass slide above the SU-8 mold. The mold cavity is flooded with low density polyacrylamide solution (8%T, 3%C, 1% w/v VA-086) and cured with UV light (13 mW/cm2) for 20 seconds. The final device can then be separated from the SU-8 mold.

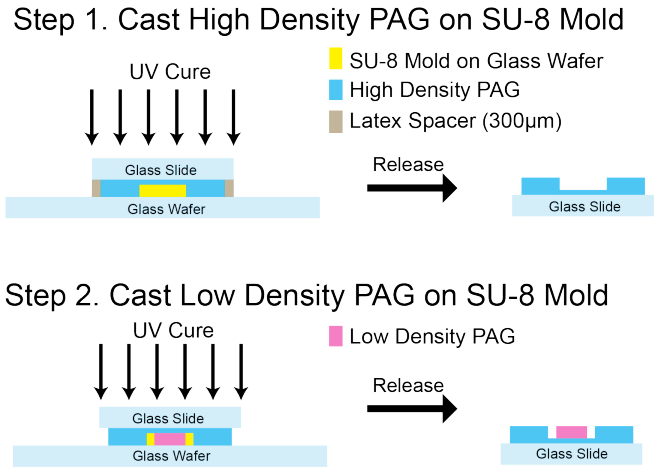


Figure 3.13: The fabrication protocol for making the glass slide based devices. The process involves two major steps. First the high density region is polymerized using an SU-8 mold to define the features. The partially fabricated device is then removed and placed onto a different mold where the low density polyacrylamide is polymerized.

The glass slide based devices have been optimized to minimize the amount of gel. Separations occur over a few millimeters in sections of gel only 250 μm thick. As will be shown in the next chapter, this makes it possible to separate DNA and RNA from small quantities of cells. Additionally the polymerization process is extremely fast, once the molds have been prepared a device can be made in roughly 20 minutes. The downside to this device is that the fabrication approach requires access to a clean room. The devices must be polymerized using a high intensity source typically found only within a clean room. Additionally, while an individual device can be fabricated relatively quickly, fabricating the SU-8 molds and

performing the various surface treatments requires a substantial investment of both time and money. A detailed protocol for fabricating the glass slide based devices can be found in Appendix B. The glass slide based device is a high performance, high cost option that may not be required for all situations.

3.3 Cassette Based Device Design and Fabrication

Many companies sell standard gel electrophoresis systems that come with a power supply, electrophoresis chamber, and empty cassettes. These systems dramatically simplify the process of conducting experiments with gel electrophoresis. End users simply fill the cassette with the desired density polyacrylamide based on their needs (recall Table 2.2). Once the gel has polymerized, they place the cassette in the electrophoresis chamber, add their sample, and use the power supply to apply an electric field. The assembled system, in this case an XCell SureLock® Mini-Cell, is shown in Figure 3.14.

Figure 3.14: The XCell SureLock® Mini-Cell electrophoresis chamber and power supply.

In an effort to make the technology presented in this dissertation more accessible to biologists and bioengineers, a cassette based device was developed. The device is shown in Figure 3.15. The left panel shows the device as it appears to the end user, the right panel shows an annotated view of the device. The functional areas of the cassette based device are the same as the glass slide based device, just with a different geometry. Samples are loaded at the top of the device into vertical wells. The first region of gel, shown in pink, is the low density membrane that permits RNA/cDNA but stops genomic DNA. The second region, shown in blue, is the high density membrane that stops the RNA/cDNA.

One substantial difference from the glass slide based device is the presence of a filler gel, shown in green. This region is required when fabricating the cassette based devices in order to create a suitable electric field throughout the device. Recall from Section **??** that high density polyacrylamide has a higher resistivity than low density polyacrylamide. If the majority of the gel cassette were filled with high density gel, then the potential difference across the low density region would be too small to induce electrophoretic migration. Therefore the filler region is composed of polyacrylamide that is the same density as the separation gel. This facilitates a mostly uniform field throughout the device. It is important to remember that the device is designed so that no nucleic acid ever enters this region of the gel.

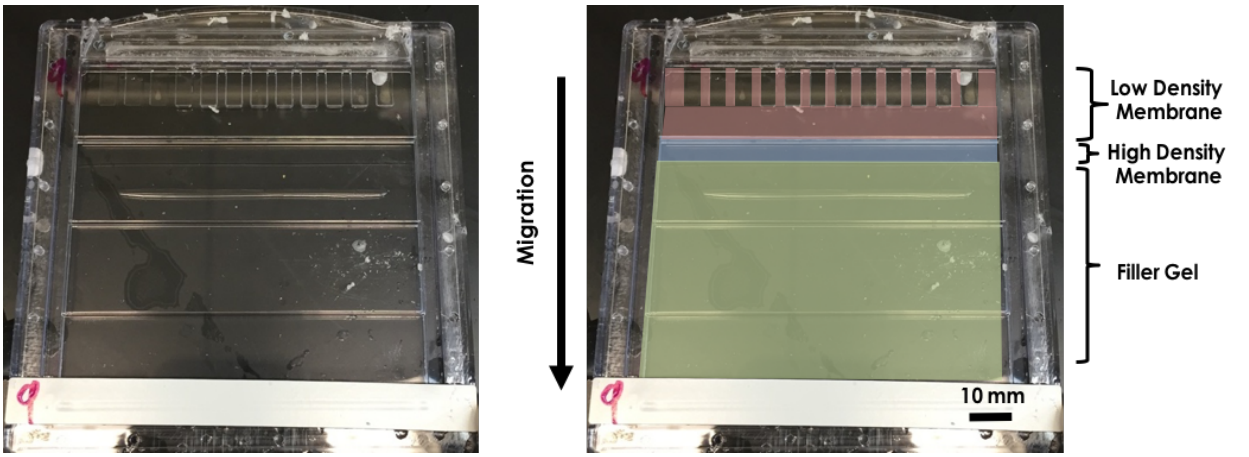


Figure 3.15: A cassette based polyacrylamide device developed for separating DNA and RNA. The left panel shows the device as it looks to the end used, the right panel has been annotated to highlight the various regions of the device.

Device fabrication is straightforward and, with the help of the detailed protocol provided in Appendix C, could be performed by researchers familiar with fabricating standard polyacrylamide gels cassettes. An overview of the protocol is shown in Figure 3.16. First the monomer solutions for each layer are made by combining the appropriate amounts of acrylamide, bis-acrylamide, TBE, water, and sucrose. (Detailed gel recipes for each of the layers can be found in Appendix C). The addition of sucrose to the polyacrylamide precursor solution facilitates the formation of smooth interface layers between the different densities, but has minimal impact on electrophoresis. The precursor solutions are vortexed to ensure thorough mixing and then degassed by applying house vacuum and immersing the tube in a sonicator. This step helps to remove dissolved gasses that inhibit the polymerization process. Just before adding the precursor solution to the cassette, the polymerization initiator APS and catalyst TEMED are added to the solution.

The layers are fabricated from bottom to the top. First, 6 mL of filler gel precursor is added to the cassette. The remainder of the cassette is then filled with de-ionized, degassed water and the filler gel is allowed to polymerize for at least one hour or up to overnight. The

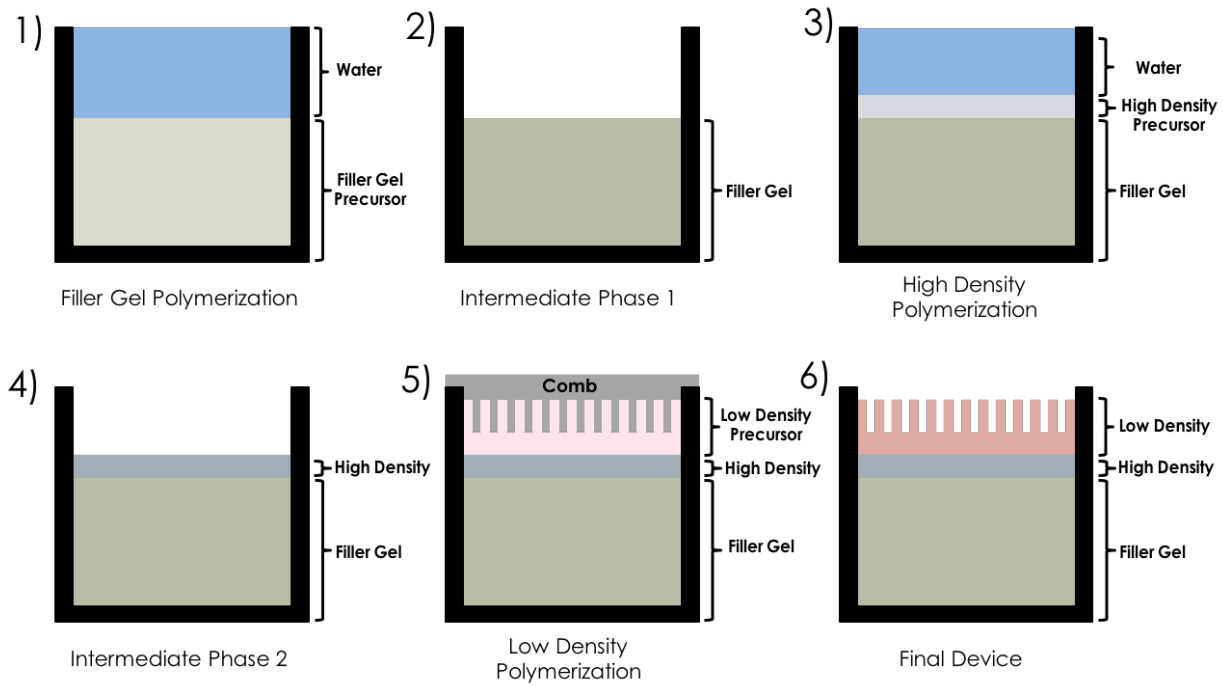


Figure 3.16: The fabrication protocol for the cassette based devices. Each layer of gel is allowed to polymerize before the next layer of gel is poured on top of it. A water overlay is helps to create a smooth interface between layers.

water overlay ensures the formation of a smooth interface. After polymerization, the water overlay is removed by inverting the cassette and shaking. Compressed air can be used to assist in the removal of any trapped water droplets. Next, 350 μL of high density precursor is added to the cassette. Due to the small volume of high density gel, it is important to make sure the precursor is evenly distributed. This can be achieved by tilting the cassette back and forth to allow the liquid to uniformly spread out over the filler gel. Once the high density precursor has been uniformly distributed, a water overlay is added. In order to obtain the best interface, it is important to add the water slowly to the center of the cassette to minimize the mixing with the high density precursor. The high density gel is allowed to polymerize for at least 10 minutes before the water overlay is removed. Finally the low density precursor is added to fill the rest of the cassette, approximately 1.65 mL. In order to define the loading wells, a template known as a comb is then inserted into the cassette. Cassettes can be fabricated with different numbers and sizes of wells by using different combs. In this work, gels were fabricated with either 10 or 12 well combs.

The cassette based devices offer several advantages over the glass slide based devices. First, the straightforward fabrication process lowers the technical challenges associated with implementing this new protocol. By taking advantage of standard equipment and slightly modifying standard polyacrylamide gel protocols, this approach lends itself to easy adoption

by those unfamiliar with fabricating devices in the cleanroom. Second, the ease of fabrication makes it possible to quickly and inexpensively test different gel chemistries. This is a key advantage during the development process, where determining the proper gel chemistry and layer thicknesses is a non-trivial step that requires multiple rounds of testing. One final advantage offered by the cassette based devices is the larger size. Larger sample volumes, up to 50 or even 100 μL can be loaded into these types of devices. Similarly, the larger size of these gels makes it easier to physically manipulate the gel when attempting to cut sections of the gel out for DNA and RNA recovery. This large size, however, is also the major drawback to this form factor. The smaller sections of gel found in the glass slide device are better for concentrating small quantities of sample and simplify the extraction of nucleic acid from the gel.

3.4 Biological Protocol Development

Recall from the introduction to genetic sequencing in Section 1.2 that the basic steps required for obtaining the genome of a biological sample are: sample preparation, amplification, library preparation, sequencing, and data processing. The Gel-Seq protocol reported on in this dissertation integrates both a physical device to separate DNA and RNA as well as the biological protocols necessary to generate genome and transcriptome data. Therefore, in parallel with the development of devices to physically separate DNA and RNA, it was also necessary to develop biological protocols compatible with these devices. This section reports on the development of these protocols. The section begins by explaining why PC3 cells were chosen as the ideal sample for validating device operation. This is followed by sub sections that discuss the protocols developed for creating sequencing libraries from both DNA and RNA.

Biological Sample Selection

The ideal biological sample would have uniform expression of DNA and RNA across the entire sample. This uniformity is important for device validation. With a uniform sample it is possible to compare genome and transcriptome libraries obtained using the device to reference libraries generated with standard protocols in tube controls. For this reason, it was decided to validate the device using a cultured cell line, PC3, rather than samples from humans or mice.

PC3 is a prostate cancer cell line. As with many cell lines, the genome and transcriptome is relatively uniform across different cells. PC3 was chosen over other cell lines because it is representative of the types of biological samples that might be investigated using this device. In particular, the PC3 genome is an example of a cancer cell with copy number variations (CNVs). As mentioned in the introduction, CNVs are extra copies of portions of the genome that have been incorporated into a cells DNA. CNVs are known to play a role in many

cancers and it is a widely studied area [36, 37]. The impact of CNVs on other diseases, ranging from Alzheimer's to Autism, is also an important area of research [38, 39].

The presence of CNVs make PC3 ideal for comparing data from the device behavior to tube controls. This can be readily seen by examining the PC3 genome, shown in Figure 3.17. This figure displays the sequencing data by grouping the raw data, known as genomic reads, into bins that span 10,000 basepairs. The x-axis shows the location of each of these bins along the entire 3 billion basepairs of the human genome. The number of reads that fall into each bin are counted, a value known as the raw bin count. As the number of reads generated for each sample can be different, the raw bin counts need to be normalized for comparison. Humans have a diploid genome, meaning that two copies of a gene are present for any particular bin. Therefore the raw bin counts can be normalized to a mean value of two. The y-axis of Figure 3.17 shows the normalized bin count at each of the bin locations. Healthy human cells would show a uniform normalized bin count of two across the entire genome. PC3, on the other hand, shows spikes and dips across the genome, indicating the presence or dropout of large regions of DNA. This unique signature facilitates the comparison of the genome obtained from tube controls to the genome obtained using the device.

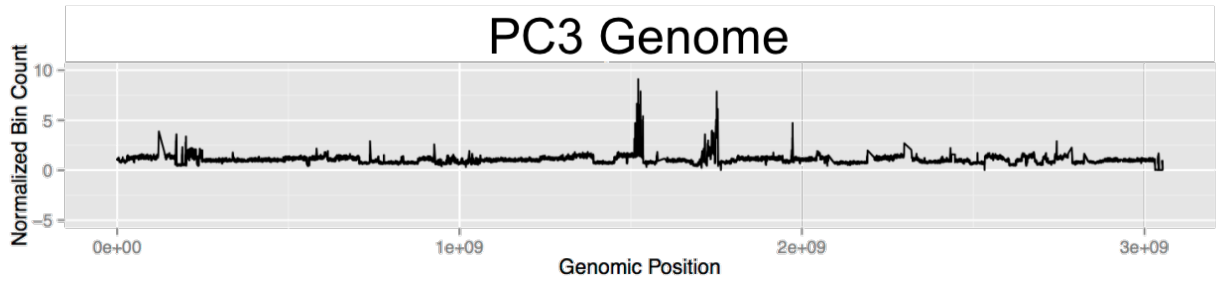


Figure 3.17: The genome of PC3, a cell line derived from prostate cancer. Healthy human cells have a diploid genome, or two copies of every gene. PC3 cells, on the other hand, contains regions with copy number variations where extra copies of certain genes have been incorporated into the genome.

Protocol for Generating DNA Libraries

When building DNA libraries, it is desirable to obtain uniform information across the genome. This has historically presented a problem when working with low input samples. While it is possible to amplify the genome, it is challenging to ensure uniform amplification that does not introduce bias [14]. As a part of their genetic sequencing workflow, Illumina has developed a streamlined low input DNA library preparation protocol known as NexteraXT [11]. This protocol requires only 1 ng of DNA, or roughly 166 cells, and has been optimized to minimize amplification bias. Research into these transposome based library preparation

protocols has shown that it is possible to further scale down NexteraXT reactions while still generating viable DNA libraries [40, 41].

In order to avoid the challenges introduced by amplification bias, the decision was made to focus on a protocol where the DNA did not need to be amplified before starting the NexteraXT protocol. Examining genome data, shown in the next chapter, it was determined that half volume NexteraXT reactions were sufficient for generating useful DNA libraries. This set a lower limit of $\tilde{1}00$ cells before amplification would be required, a sufficiently low input for many applications.

A detailed NexteraXT protocol can be found on the Illumina website [11]. The only modification made to this protocol was to reduce the reaction volumes by half. In brief, the protocol consists of two major steps, transposition and PCR. In transposition, an enzyme known as a transposome is used to fragment the DNA into sequences ranging from 200 - 800 base pairs in length. (This is the optimal size for sequencing with Illumina machines.) During this step the transposome also adds sequencing adapters to each end of the DNA fragment that indicate the start and end of the fragment during sequencing. These sequencing adapters are then used in the PCR step to add additional sequencing primers with unique index numbers during the first cycle of PCR. The fragments are amplified with an additional 10 - 12 cycles of PCR, producing sufficient product to be used for sequencing. As these fragments are all the same size, there is little amplification bias introduced by this step.

The NexteraXT protocol calls for input DNA to be dissolved in molecular-grade water or 10 mM Tris HCl (pH 7.58.5) at a concentration of 0.2 ng/μL. Additionally, the input DNA must be double stranded for the transposition enzyme to fragment the DNA. These constraints will be revisited when discussing the extraction protocol during the next chapter. They also play a role in the decisions made about the protocol for generating RNA libraries.

Protocol for Generating RNA Libraries

A major breakthrough in transcriptomics was the development of a protocol known as RNA-Seq [42]. This protocol laid out the basic foundation for using next generation sequencing methodologies to quantify the transcriptome. The process for generating low input RNA libraries is substantially more challenging than the process for generating DNA libraries. While there are roughly equivalent amounts of DNA and RNA contained within a cell, mRNA makes up only 1 - 10% of total RNA [5]. Recall from Section 1.1 that mRNA is the primary molecule of interest when examining gene expression, which means that in 100 cells there could be as little as 6 picograms of relevant input material. These vanishingly small quantities of RNA present the fundamental challenge for generating low input transcriptome libraries. In the case of low input RNA, some amplification is required before sequencing.

The first step in generating an RNA library is to convert the RNA into cDNA. With sufficient amounts of cDNA, it is possible to use the NexteraXT protocol described in the previous section to generate sequencing libraries. Therefore they key to generating low input

RNA libraries is to find a reliable and uniform method for 1) converting RNA to cDNA and 2) amplifying the cDNA by several orders of magnitude.

Since the introduction of RNA-Seq in 2009, many derivative protocols have been developed [43, 15, 17]. The current gold standard for RNA transcriptomics is a protocol known as Smart Seq [17]. While this protocol does an excellent job of producing high quality cDNA libraries from mRNA, it is also extremely expensive. Reagents for a single reaction are roughly $100, making it cost prohibitive during the initial development stage of a new technology where a single experiment requires 5 to 10 reactions. Therefore a similar but less expensive protocol, CellAmp, was selected for the research and development phase. The protocol, originally developed by Kurimoto [16], is sold by Clontech as the CellAmp Whole Transcriptome Amplification Kit. The reagents for this approach cost a more reasonable $4 per reaction.

CellAmp has been designed for low starting inputs, just 1 - 1000 cells. The protocol has been optimized to both convert the RNA to cDNA and then amplify the cDNA from picogram to nanogram quantities. The first step is to lyse the cells using a gentle detergent lysis. This is followed by reverse transcription, where the mRNA is converted to cDNA. One unique aspect to the reverse transcription step in this protocol is that only the first 1000 nucleotides of each transcript are converted to cDNA. (This is achieved by limiting the reverse transcription reaction time to only 5 minutes.) By limiting the reverse transcription, the cDNA transcripts are all relatively similar in size. This helps to limit amplification bias in the downstream steps.

Once the cDNA has been synthesized, a poly-A tail is added to the cDNA molecule using the Terminal deoxynucleotidyl transferase (TdT) enzyme. This creates cDNA molecules with defined sequences on either end, AAAAA... at one end and TTTTT... at the other. The last step in the protocol is to amplify the cDNA molecules using PCR. Rather than using the entire RT product, which would inhibit the PCR reaction, about $\tilde{2}0\%$ of the product is used as input for PCR. The poly-A and poly-T sequences on either end of the cDNA molecule are used to attach unique primers for PCR, ensuring that there is a relatively uniform amplification across all of the transcripts. After amplification, the final step is to prepare the cDNA for sequencing using NexteraXT. The protocol used is identical to the one described in the previous section.

Development of Gel-Seq Protocol

As mentioned in the introduction, there are two approaches that can be taken when separating genomic DNA and RNA. Option one is to separate the RNA from the genomic DNA. Option two is to convert the RNA to cDNA and then separate the cDNA from the genomic DNA. While it is possible to maintain the sterile conditions needed to preserve RNA integrity during separation, it is preferable to convert the RNA to cDNA as soon as practically possible. The sooner the RNA is converted to cDNA, the less likely that the information will be lost due to RNA degradation. Therefore it is desirable to develop

a protocol that converts the RNA to cDNA before separation, but does not damage the genomic DNA.

Examining the cell amp protocol, the only step that would damage the genomic DNA is PCR. PCR requires repeated temperature cycling with a denaturing step at 98°C. This step denatures the cDNA, enabling the synthesis of two new cDNA molecules based on the previous copies. The problem with this step is that it would also denature the genomic DNA, leaving it single stranded. (Recall that NexteraXT protocol for preparing the genomic DNA libraries requires double stranded DNA.)

With this insight into the biological implications, a combined Gel-Seq protocol based on the CellAmp Protocol and NexteraXT was developed. (A step by step protocol is provided in Appendix D The protocol workflow is shown in Figure 3.18. The protocol starts with 1 - 1000 cells that have been washed and suspended in PBS. The cells are then added to a lysis buffer (final concentration: 50 mM KCL, 10 mM Tris-HCl, 0.5 % NP-40, 5 mM DTT, and 10 units RNAse inhibitor). To ensure complete lysis, the tubes are heated to 70°C for 90 seconds. This lysis makes the genomic DNA (gDNA) and RNA accessible for further processing. Next, the RNA is converted to cDNA. Reverse transcription (RT) enzymes are added and the tubes heated to 42°C for 5 minutes. To inactivate the RT enzyme, the tubes are then immediately heated to 85°C for 10 seconds. In order to remove the unused single stranded DNA primers, which can over-amplify during PCR and swamp the relevant signal, Exonuclease I is added to the samples. The tubes are then incubated at 37°C for 15 minutes, allowing the Exonuclease to break down the primers, and then the reaction is stopped by incubating at 80°C for 15 minutes. After removing the unwanted primers, poly-A tails are added to the newly synthesized cDNA using the TdT enzyme.

At this point in the protocol, the sample is prepared for the physical separation. The RNA has been converted to the more stable cDNA, while the gDNA remains intact as long, double stranded fragments. The sample, $\tilde{1}0$ μL, is mixed with 8 μL of loading dye and added to the device. The gDNA and cDNA are then separated by applying an electric field of $\tilde{2}5$ V/cm. The nucleic acids are then recovered from the device, the details of this protocol will be discussed in the next chapter. The recovered gDNA is used as input for the NexteraXT protocol to generate a DNA library. In parallel, the cDNA is amplified using PCR and an RNA derived library is generated using the NexteraXT protocol.

3.5 Chapter Summary

This chapter presented the foundational work in the development of the Gel-Seq protocol. In order to confirm the major design choices, the validation process was broken out into a series of discrete tasks. Based on the insights gained from these initial experiments, two devices were developed: a glass slide based device and a cassette based device. The glass slide based devices are designed to minimize the required sample input by utilizing thin layers of polyacrylamide. These thin layers require the use of a clean room to fabricate both molds and the devices themselves. The cassette based devices, on the other hand, are designed

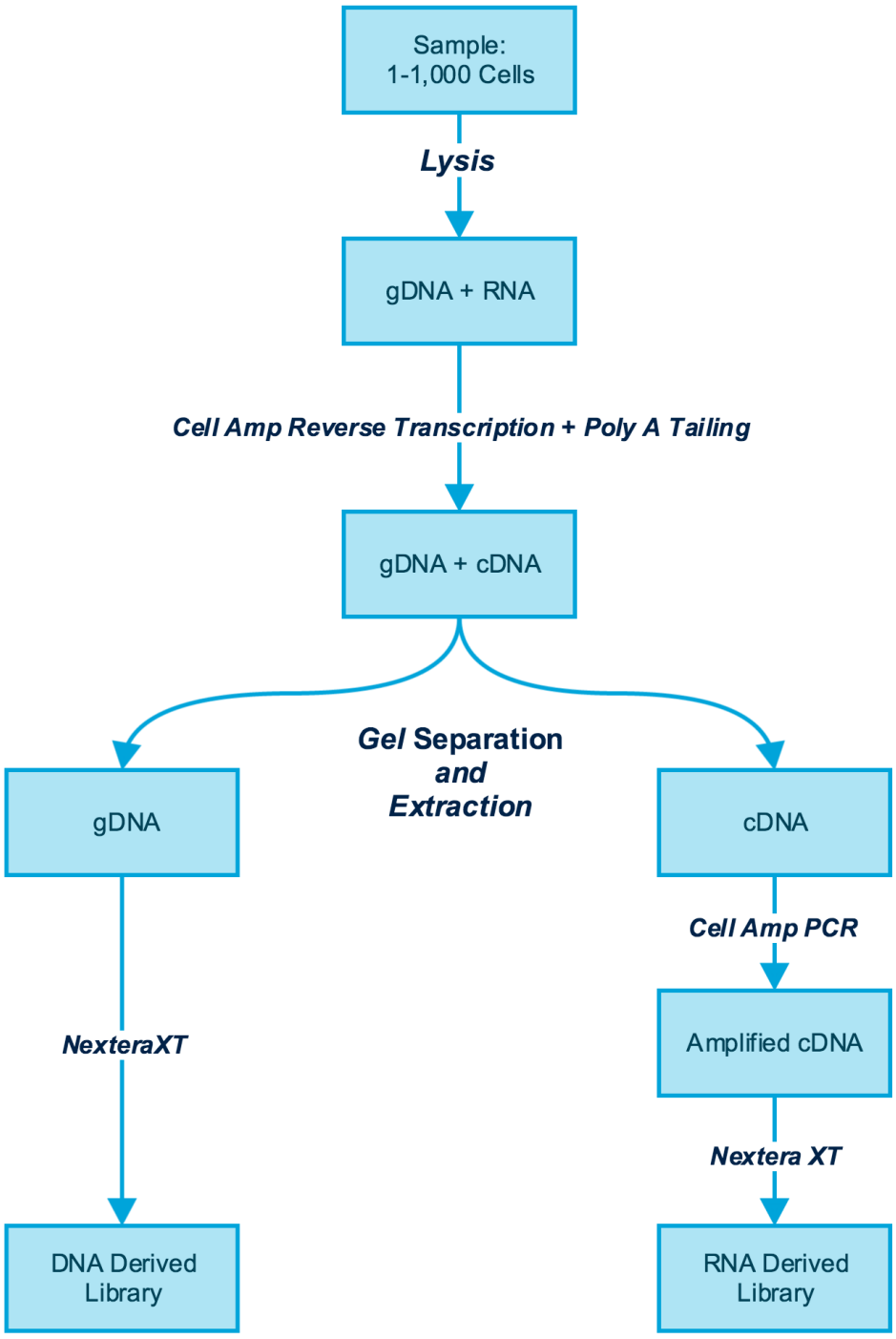


Figure 3.18: An overview of the Gel-Seq protocol for generating matched DNA and RNA libraries for sequencing.

for higher inputs as well as easy adoption by the biology community. These devices can be fabricated and tested using standard equipment already present in many biology labs. In addition to the development of physical devices, it was also necessary to develop biological protocols that were compatible with a physical separation. The Gel-Seq protocol draws from two protocols: NexteraXT and Cell Amp. This protocol converts mRNA to cDNA and then uses the devices to physically separate the gDNA from the cDNA. Once separated, the samples are further processed to create matched genome and transcriptome libraries.

Chapter 4

Optimization and Validation

The preliminary validation experiments presented in the last chapter suggested that polyacrylamide devices could be used to separate DNA and RNA. This separation, however, is only one component of building a device to sequence both the genome and transcriptome. This chapter reports on the optimization of the design and overall validation of the technology. The cassette based Gel-Seq device is best suited for widespread adoption, therefore primary emphasis has been placed on validating this form factor. The chapter begins by reporting on the process of tuning the cassette based gel chemistry. The different gel chemistries are evaluated using fluorescent imaging. While fluorescent imaging is useful for preliminary validation, ultimately the device performance must be confirmed with sequencing data. In order to generate sequencing data, the sample must be recovered from the device. The second section of this chapter discusses the development of a sample recovery protocol and presents the steps used to build genomic libraries from these recovered samples. This sequencing data, in comparison with tube controls, is used to evaluate device performance. In the last section, preliminary experiments to evaluate the glass slide based device performance are presented.

4.1 Separation Optimization

Multiple rounds of testing were required to determine the optimal chemistry for separating DNA and RNA/cDNA. This optimization required balancing several factors. First, the two membrane chemistries must serve their intended function: one membrane must stop genomic DNA but permit RNA/cDNA, the other membrane must stop RNA/cDNA but permit buffer ions. In addition to the constraints based on the primary design function, there are also physical limits to the types of gels that can be fabricated. Below 3.5%T, gels take on a mucus like consistency and fall apart when removed from the cassette. Conversely, it is impossible to fabricate gels above $\tilde{5}0$%T due to the solubility limits of both acrylamide and bis-acrylamide. The thickness of the high density membrane had an effect on the gel properties. Adding more than 1 mL of high density membrane would often cause residual

stress in the gel that lead to delamination of the gel from the cassette.

The interface between the high and low density membranes is also an important area for consideration. For example, attempting to fabricate a 3.5%T gel on top of a thick 50%T gel often caused the gels to separate at the interface. This was likely due to the fact that during polymerization the 3.5%T gel was not able penetrate the 50%T gel, preventing the chain entanglement that held the interface together at other gel chemistries.

The results from a range of tests with different gel chemistries are shown in Figure 4.1. These images, similar to those shown in Section 3.1, are fluorescent images that show the location of DNA and RNA ladders that have been stained with SYBR Gold. The samples are loaded into the wells of the device and an electric field was applied using the XCell SureLock® Mini-Cell electrophoresis chamber (shown previous in Figure 3.14) and an external power supply. In some tests, such as Panel A, the high density gel did not stop the ladder. In other tests, such as Panel B and D, there was insufficient migration as most of the ladder remained near the loading well. Panel C shows an example where one of the ladders (DNA in this case) piled up at the interface, but the RNA ladder remained spread throughout the separation gel. Panel E shows an example where filler gel ripped. There are other similar cases where the separation and high density gel ripped apart, though they were not imaged because they had disintegrated.

After several rounds of experimentation, the optimal gel configuration was determined to be a 350 μL 40%T, 5%C high density region beneath a 1.7 mL 4%T, 3.3%C low density region. The filler gel was the same density as the low density region. The results from an experiment with this gel chemistry are shown in Figure 4.2. In this case the gel was run for 60 minutes with an applied voltage of 200 V, producing an average electric field of 25 V/cm. (Recall from Section 3.1 that there is actually not a uniform field throughout devices with different density layers of polyacrylamide, so the applied voltage is the most important parameter.)

The figure shows four different types of separations. Notice first lanes 1 and 7, which were loaded only with water. These negative controls are useful references for quantifying the background autofluorescence present in the gel. (In this experiment there was minimal autofluorescence, which is shown by the lack of black bands present in these lanes.) Next consider lanes 2 and 8, which show the behavior of a DNA ladder with fragments ranging in size from 100 - 2000 bp. These lanes show that the ladder has migrated through the entire separation gel and been completely stopped at the interface between the high and low density membranes. The samples loaded into lanes 3 - 6 demonstrate the separation of genomic DNA and cDNA. Before separation, these samples were processed to convert the RNA into cDNA as described in Section 3.4. The black band at the top of the low density membrane indicates the presence of genomic DNA that has been trapped. The band present at the interface with the high density membrane, fainter than that of the genomic DNA, indicates that the cDNA has been trapped as well. Note that this band is likely a mix of both cDNA and RNA, as the RNA is not removed from the sample during reverse transcription.

In addition to separating gDNA and cDNA, this experiment also suggests that gDNA

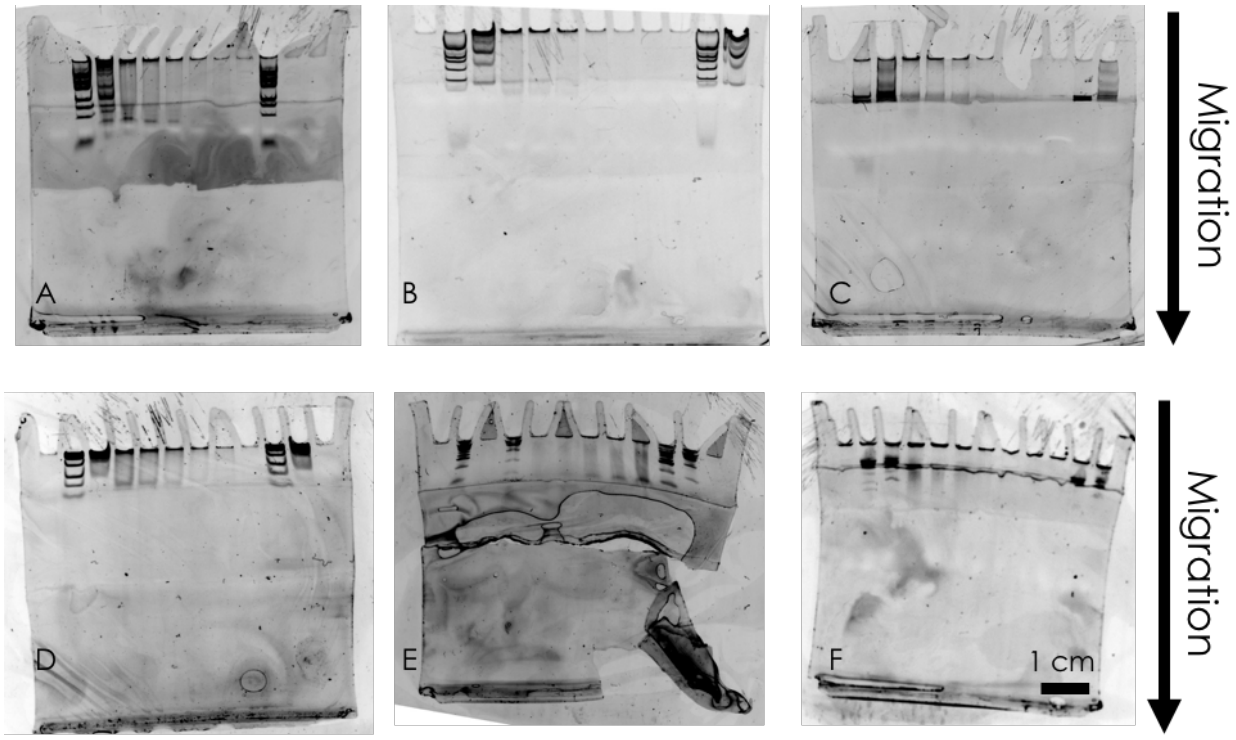


Figure 4.1: Experimental results from testing multiple gel chemistries. Multiple rounds of optimization were required before the ideal gel chemistry was identified.

can be separated from RNA (lanes 9 - 12). In this case, cells were lysed open using the same detergent lysis in the first step of the standard Gel-Seq protocol. Rather than converting the RNA to cDNA, the sample was kept on ice to preserve the RNA and then loaded into the device along with the remainder of the samples for separation. The black bands present at the interface between the high and low density regions indicate that the RNA has been separated from the gDNA.

This experiment suggests that it is possible to separate DNA from RNA/cDNA, however this conclusion is based exclusively on imaging data. These data are not sufficient to confirm this hypothesis. While a ladder was used as a control and indicates that the gel seems to be achieving the desired separation based on fragment size, it is possible that some RNA/cDNA is being trapped by the genomic DNA and not moving through the low density membrane. Similarly some of the genomic DNA could have degraded and passed through the low density gel, masquerading as RNA/cDNA. To fully validate the device, samples must be extracted and compared to tube controls through sequencing.

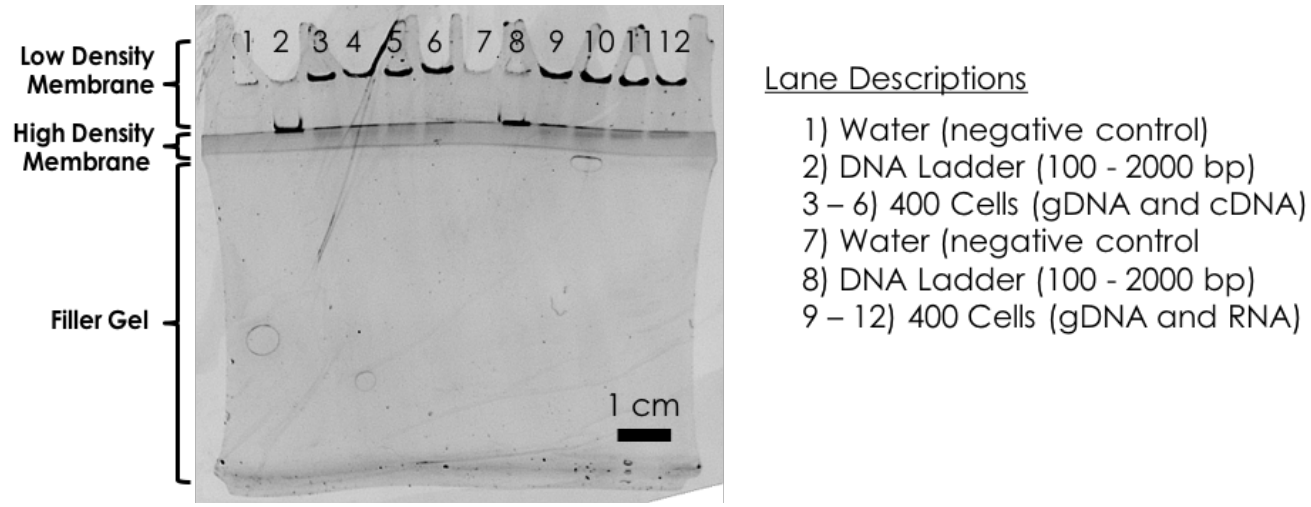


Figure 4.2: The results from a cassette based separation with optimal gel properties. This experiment shows the separation of a DNA ladder, cells processed with Cell Amp, and cells lysed but otherwise unprocessed.

4.2 Sample Recovery and Library Construction

The recovery of DNA and RNA from polyacrylamide gels is a challenging problem with no standard solution. A comprehensive review article by Kurien and Scofield cites over 75 papers and describes the multitude of approaches developed by researchers [44]. The novel gel membranes developed for this device, as well as the nature of the samples to be recovered, required the development of a unique recovery approach. Two samples must be recovered, long genomic DNA molecules from 4%T gel and shorter RNA/cDNA from the interface of a 4%T and 40%T gel. The limited amount of sample, on the order of nanograms, eliminated many of the common kits sold by manufactures for the recovery of nucleic acid from hydrogels. Several different recovery methods were tested, including the creation of dissolvable polyacrylamide gels. In the end, the most successful approach was based on the tried and true "crush and soak" method. (A standard crush and soak protocol has been published by Cold Spring Harbor, though there are multiple versions [45].)

At its heart, the crush and soak approach relies on diffusion to extract DNA or RNA from polyacrylamide. The first step in this protocol is to cut out the small section of the gel containing the gDNA or cDNA/RNA. Typically, a UV light box is required to visualize the sample during the cutting stage. An advantage to the Gel-Seq devices is that the sample location is clearly defined and visualization is not necessarily required during sample recovery. The gDNA is trapped at the top of the separation gel, while the RNA/cDNA is located at the interface between the two gels. Using a clean scalpel it is relatively straightforward to cut small sections of gel, roughly 5mm x 2 mm x 1 mm, from the device. Note that when recovering the RNA/cDNA samples, the sample was cut to include regions of both the high

and low density membrane. Each gel sample is then placed into a separate tube using sterile tweezers and pulverized using a pipette tip. The high density gel is relatively brittle and quickly breaks apart, while the low density gel requires more careful attention to ensure uniform breakup. This crushing step is integral for facilitating diffusion: by creating small fragments of gel the diffusion length is dramatically shortened and the sample can more easily escape from the gel.

After pulverization, water is added to the sample. Typically DNA and RNA are eluted into a buffer, however in this case water was chosen to facilitate a downstream step calling for the reduction of volume. For the DNA samples, where the starting volume for the NexteraXT reaction needed to be minimized, just 20 μL of water was added to each sample. For the RNA/cDNA samples, where the downstream reaction could tolerate a larger input volume, 80 μL of water was added. The tubes are then incubated at 37°C for 12-16 hours. During incubation, the tube is shaken using a vortex mixer. This constant mixing, combined with the elevated temperature, further encourages the diffusion of the nucleic acid out of the gel. After removal, the tubes are gently spun in a centrifuge to collect the gel fragments at the bottom of the tube. The supernatant containing the eluted gDNA or RNA/cDNA is transferred via pipette to another tube.

After separation, processing proceeds in two parallel paths. The eluted volumes containing gDNA need to be concentrated to 0.5 ng/μL, the input concentration for NexteraXT. Depending on the sample, between 2 and 10 ng of gDNA is recovered from the device, which is then diluted into 20 μL of water. This requires that the sample be concentrated before it can be used as input for the NexteraXT protocol. In this case, the sample concentration can be increased using a Vacufuge Concentrator. Depending on the volume and concentration, samples are spun for 20 - 40 minutes at 35°C to reduce the volume.

A vacufuge resembles a standard centrifuge, however there is a vacuum in the centrifuge chamber. By spinning open tubes inside of this tool, the sample volume is decreased due to rapid evaporation. The tubes are constantly spun to ensure that the sample does not adhere to the walls. While vacufuges are an efficient way to increase sample concentration, they only remove water from a sample. This means that if a sample is suspended in buffer, using a vacufuge would also increase the salt concentration. Substantial reductions in volume lead to high salt concentrations that can inhibit downstream reactions. This is why the gDNA is eluted into water, enabling the use of a vacufuge without concern for increasing the concentration of other species. Once the gDNA is sufficiently concentrated, a gDNA derived library can be generated following the NexteraXT protocol.

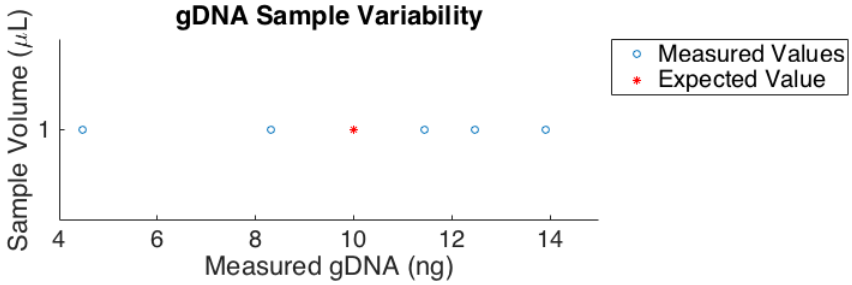
For the samples containing cDNA, the next step is to perform PCR amplification. Recall from Section 3.4 that only $\tilde{2}0\%$ of the RT product is required as input for the PCR step. As the cDNA has been eluted in water, it is possible to setup an equivalent PCR using 12 μL of the diluted cDNA recovered from the gel. Once this cDNA is amplified, an RNA derived library can be generated using NexteraXT.

The literature for the crush and soak method report yields ranging from 50%-90% [44]. Unfortunately quantifying the yield for the Gel-Seq device is extremely challenging. While it is possible to quantify the amount of gDNA present in small samples quite accurately, the

measurement process consumes the majority of the sample. This means that quantifications can only be performed either before or after separation, but not both. Unfortunately, when working with very dilute concentrations of cells, there is substantial variability between samples. This variability means that controls cannot be used to provide a reference starting concentration against which the recovered sample from the gel could be compared.

The variability between samples was confirmed with a small experiment. Five samples were prepared where 1 μL of a 1500 cells / μL cell solution was lysed by adding 3 μL of lysis buffer. The gDNA content for each sample was then measured using a Qubit fluorometer assay. The expected gDNA content for 1500 cells is roughly 10 ng. The results from the quantification, shown in Figure 4.3, show large variability in the measured gDNA content between the samples. This variability is not due to differences in the gDNA content of the cells, which are extremely minor. Most likely the differences are due to errors associated with pipetting small volumes and/or non-uniform cell lysis.

Figure 4.3: Results from a repeatability test quantifying sample variability.



While it is possible that an experiment could be developed to quantify the device yield, maximizing yield is not a design goal. As long as sufficient material is recovered to create robust sequencing libraries, then the device has functioned properly. Therefore the real test is to compare libraries generated from the device to tube controls.

4.3 Sequencing Validation

While preliminary experiments and imaging suggested the device was properly separating gDNA and RNA/cDNA, ultimately the device must be validated using sequencing data. Recall from Section 3.4 that a cell line, PC3, was chosen because it has consistent DNA and RNA expression between cells. Using a stable cell line allows for the comparison between data generated in tube controls, processed using standard protocols, and data generated using the Gel-Seq protocol. This subsection presents validation results for both the genome and transcriptome data.

In order to minimize any variability between samples due to degradation, validation experiments were conducted using PC3 cells from the same frozen aliquot. The sample was thawed and immediately processed in three different ways. All three processes were run in parallel to minimize the potential for sample degradation. One set of samples was

processed using the Gel-Seq protocol previously described, generating both genome and transcriptome libraries. Another set of samples was processed in tube to generate genome libraries. The cells were lysed using a mild detergent and the gDNA was used as input for the standard NexteraXT protocol. The third set of samples were processed in tube to generate a transcriptome library. These samples were processed by following the standard CellAmp protocol followed by the standard NexteraXT protocol. All of the libraries were then sequenced using an Illumina Miseq, generating data that could be used to compare between the tube controls and Gel-Seq protocol.

Genome Validation

The most straightforward way to compare the tube and Gel-Seq genome samples is to examine plots of the normalized bin count. (This type of plot was previously introduced in Section 3.4, see that section for more details about this presentation format.) Figure 4.4 shows the data from the Gel-Seq device as well as two different tube controls. A visual comparison of these plots shows strong agreement between the two methods.

This comparison can be made more rigorously using a Pearson correlation, often referred to as the correlation coefficient or simply "R". This approach measures the linear relationship between two samples. The correlation between two data sets, X and Y, can be calculated as follows [46]:

$$R = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2}\sqrt{\sum (Y_i - \overline{Y})^2}} \tag{4.1}$$

where $\overline{X}$ and $\overline{Y}$ are the averages of the data sets. This test, shown in Figure 4.5, compares the bin count at each bin location between two samples. The ideal case, comparing a sample to itself, is shown in the top figure. In this case the samples are completely correlated and have an $R = 1$. The next sub plot shows the correlation between the two tube controls. While these two tube controls were treated in an identical manner, the correlation shows that the data is not identical. This is not unexpected as there are stochastic variations in both how the libraries are generated and how they are sequenced. This comparison represents a best case that could be achieved with the Gel-Seq protocol.

Keeping these stochastic variations in mind, the data reveal a very strong correlation between the tube control and the Gel-Seq data. Considering this correlation, in conjunction with the visual review from Figure 4.4, it appears that no substantial information is being lost due to the Gel-Seq protocol. It is therefore reasonable to conclude that the Gel-Seq protocol maintains the integrity of the genomic DNA and facilitates the creation of high quality libraries for sequencing.

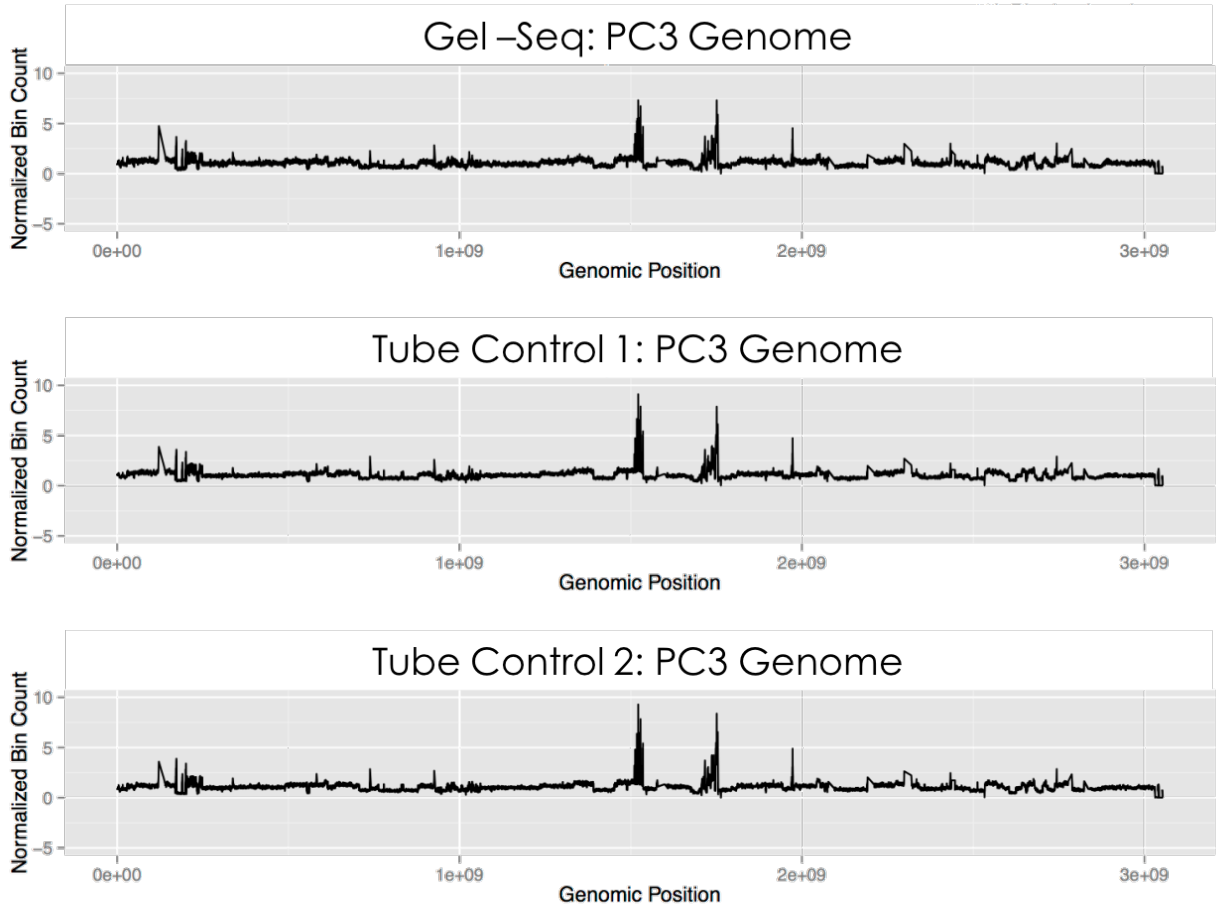


Figure 4.4: The PC3 Genome generated using the Gel-Seq protocol compared to two tube controls. Data was generated from 400 cells, using a 10,000 bp bin size and 4 million genomic reads.

Transcriptome Validation

Comparing transcriptome data is slightly more complicated than comparing genome data. The data generated from the transcriptome libraries produce reads that correspond to the RNA transcripts / genes expressed by the cell. The standard approach for interpreting RNA data is to build a gene expression list by mapping each read to its corresponding gene. These lists must then be normalized, however unlike genomic data there is no expected expression to assist with normalization. For protocols like Cell Amp, which have uniform amplification across all gene lengths, samples can be normalized using simply the total number of reads generated. This normalization makes it possible to compare samples across different sequencing runs. In other protocols, where PCR amplification bias skews the transcript counts towards longer transcripts, normalization becomes more complicated [47].

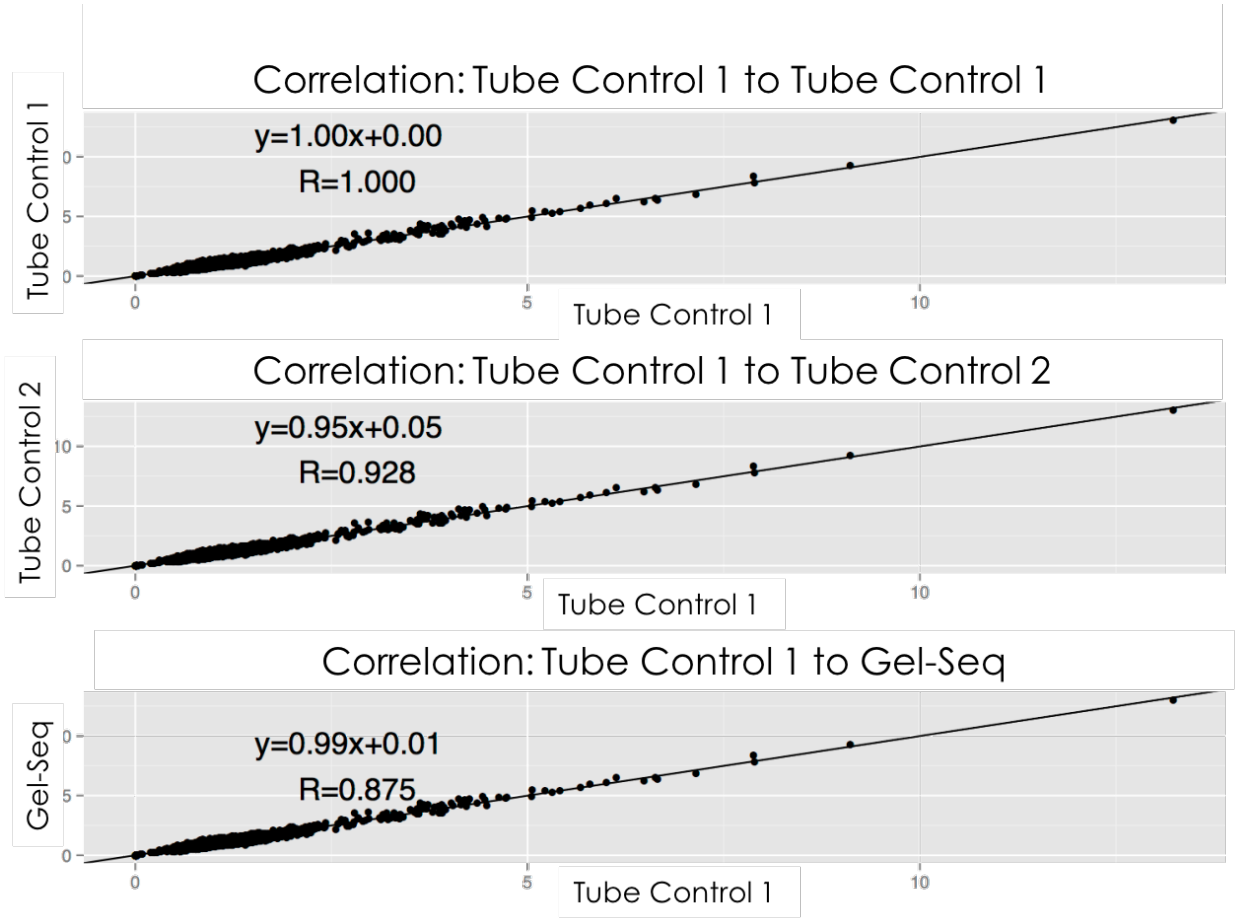
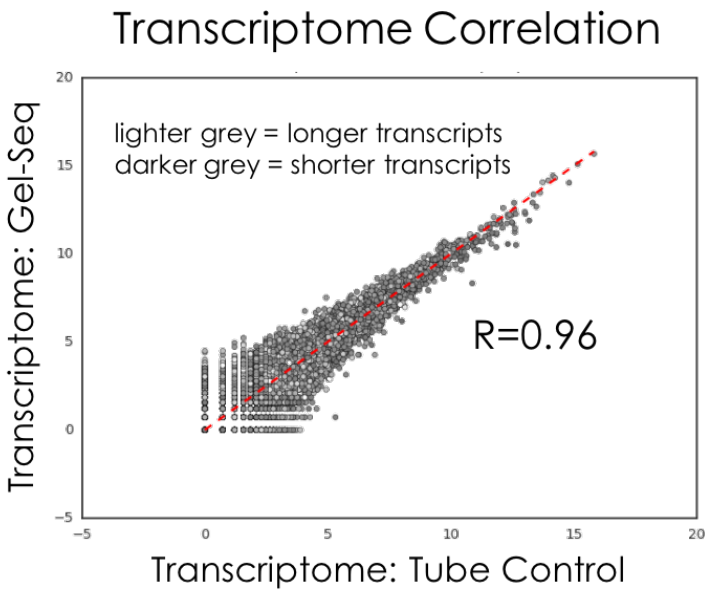


Figure 4.5: Pearson correlations comparing the genome data from a tube control to the Gel-Seq protocol. The top panel shows the reference compared to itself, a perfect correlation. The center panel shows a comparison between two tube controls treated in identical fashion, indicating the variations found between normal samples. The bottom panel shows the correlation between the Gel-Seq protocol and the tube control.

After the gene lists have been generated and normalized, they can be correlated between different samples. A correlation between a tube control and transcriptome generated using the Gel-Seq protocol is shown in Figure 4.6. This figure, which includes a table showing data for several other replicate experiments, shows a strong correlation between the Gel-Seq protocol and the tube controls. Note that for plotting purposes, the transcript counts have been log normalized to simplify data visualization. On the plot, each dot represents a particular transcript. The color of the dot has been scaled relative to the length of the transcript to investigate if the data contains any size bias. The seemingly random scattering of colors suggests there is no such size bias. These results indicate that the Gel-Seq protocol faithfully reproduces accurate transcriptome information.

As an aside for readers unfamiliar with RNA-seq data, the reader will notice several rows of dots present at 0, 1, 2 on the X and Y axes. These are the rarest transcripts and, due to the stochastic variations associated with library preparation and sequencing, are not always captured in both data sets. The rows correspond to transcripts that were present in one sample, but not found in the other. This pattern is commonly seen in transcriptome data generated using protocols in the RNA-seq family.



Replicates	R Value
A	0.85
B	0.92
C	0.95
D	0.96

Figure 4.6: Correlation of the transcriptome data generated in tube and with the Gel-Seq protocol. The figure on the left shows two gene counts plotted against each other for a Pearson correlation. The table on the right shows the correlation values for additional experiments, demonstrating reproducibility.

4.4 Glass Slide Based Device Optimization and Validation

The glass slide based devices are designed to allow processing of rare samples. While the fabrication protocol is more complicated than the cassette based device, the fundamental principles of operation are the same. Therefore the validation steps previously described can be similarly applied to the glass slide based technology.

Similar to the cassette based device, a major challenge was the identification of the optimal membrane chemistry and electrophoresis conditions for separation. Recall that the fabrication protocol for the glass slide based devices is very similar to the fabrication protocol used in the initial validation experiments. Therefore many of the lessons learned in these experiments were useful in helping to select the optimal membrane chemistry. The results of four experiments are shown in Figure 4.7. Panels A and B show results from previous generations of devices used during the preliminary optimization phase. Migration was induced from left to right in these samples. In both panels A and B, the RNA/cDNA ladder moves only partially through the separation gel. This indicated that the separation membrane was too dense, insufficient time was given for the polymer to move through the gel, and/or the applied electric field was too low. Panels C and D show experiments using the final device layout, migration is from top to bottom. In panel C the low density membrane was too dense and nothing was able to enter the gel, instead the ladders, RNA, and gDNA all pile up at the interface of the loading well. Panel D shows a case where the electrode wicks were not parallel. This created an electric field that was not properly aligned with the well placement, leading to the off target migration of the DNA ladder. In addition, like panels A and B, the ladder did not fully move through the separation gel.

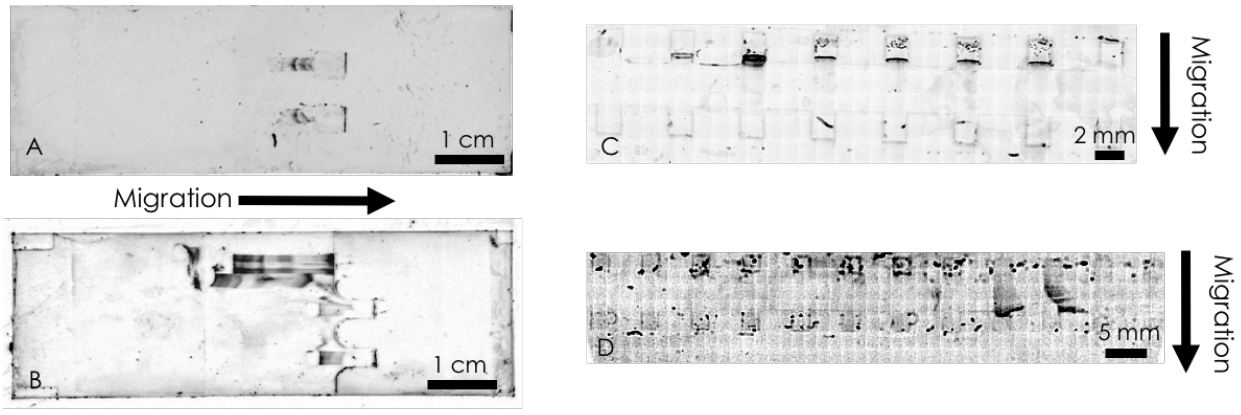


Figure 4.7: Results from four experiments optimizing the gel chemistry for separation in the glass slide based device.

After several rounds of experiments, an optimal gel chemistry for the separation was determined. The high density membrane is composed of 40%T, 5%C precursor with 1% w/v

VA-086 initiator. This precursor mixture is polymerized with 45 seconds of exposure to a 13 mW/cm^2 UV source. The low density region is made from a precursor containing of 8%T, 3%C exposed to the same UV source for only 20 seconds. The reader will note that these are different gel densities than those used for the cassette based device. The polymerization process in the two protocols is quite different and, therefore, requires different precursor concentrations to achieve the same result.

Similar to the cassette based device, fluorescent images were used to validate device performance. The device was loaded with three types of samples: DNA ladders, RNA ladders, and lysate from 1000 cells. Note that the lysate contained RNA not processed cDNA. This is because the loading wells only can only accept 1 μL of sample and the generation of cDNA increases the sample volume.

After loading, an electric field of 30 V/cm was then applied for 20 minutes. The results are shown in Figure 4.8. The resulting fluorescent image shows the expected behavior. The lanes loaded with ladder show crisp black bands only in the capture wells. The lanes loaded with 1000 cells, however, show two bands: one of gDNA in the loading well and one of RNA in the capture well. This result, combined with the data gathered from validating the cassette based device, suggested that the glass slide device has been properly tuned for separating DNA and RNA.

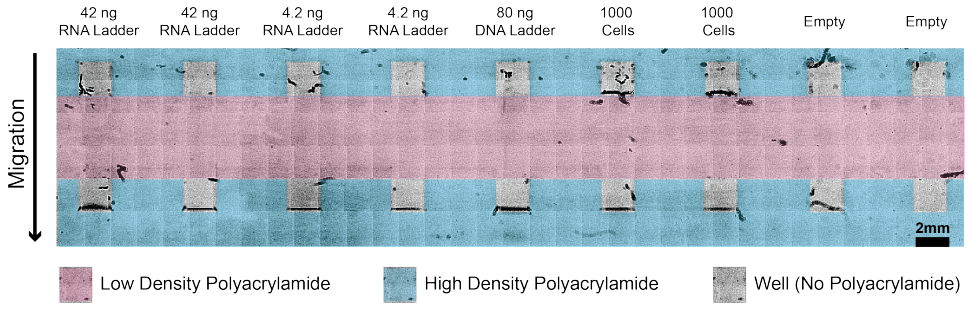


Figure 4.8: Results showing the separation of gDNA and RNA in 1000 cells using the glass slide based device.

The glass slide device is intended for testing rare samples, therefore the next phase of optimization was to examine the impact of reducing sample input. Before making measurements of the separation device, however, it was necessary to quantify the sensitivity of the fluorescent measurement technique. A special device was fabricated to quantify the detection limit of the fluorescent measurements. The device contains a row of loading wells surrounded entirely by high density polyacrylamide gel. This layout provides the standard condition for imaging, DNA on the edge of a high density membrane, while eliminating the low density region, where sample may become inadvertently trapped. The wells were loaded with different concentrations of DNA ladder, ranging from 0 to 30 ng, and an electric field of 30 V/cm was applied for 15 min. This electrophoresis step forced the DNA out of solution

and deposited it on the wall of the well. After 30 minutes of staining in SYBR gold, a fluorescent image was captured, see Figure 4.9.

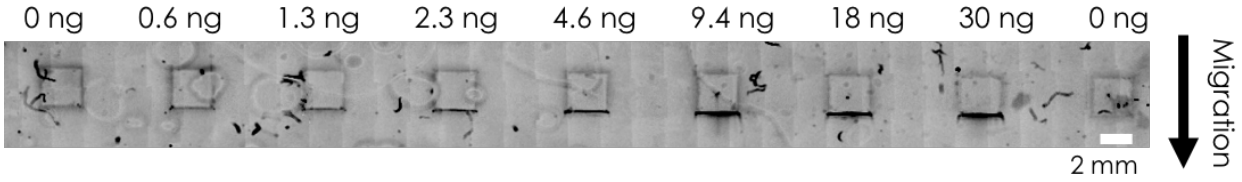


Figure 4.9: A fluorescent image showing the detection limit of the measurement techniques.

Examining the results of this experiment, it is possible to determine the lowest concentration of DNA that can be resolved using this setup for the fluorescent measurement technique. The DNA ladder is clearly present when 30 ng is loaded into the well. When 0.6 ng is loaded, however, the results are nearly indistinguishable from a negative control. The well loaded with 2.3 ng of DNA ladder has a distinguishable band, while the well loaded with 1.3 ng is at the edge of the detection limit. Recalling that each cell has roughly 6 pg of gDNA, this experiment indicates that the detection limit lies somewhere between 200 and 400 cells.

This experiment was useful for providing a rough guideline for the optical limit of detection, however it should not viewed as a rigorous analysis. There are a number of factors that impact the lower limit of detection. For instance, devices with the same chemistry have been observed to have different background autofluorescence at the membrane interface. This, coupled with the image exposure settings, influences the detection limit. (If there is sufficiently low background, it is possible to take longer exposures thereby increasing the signal to noise ratio and resolving smaller amounts of DNA.)

Understanding the metrology constraints, it was possible to test a separation close to the optical detection limit: 250 cells. The high and low density membranes were fabricated in the same manner as the device shown in Figure 4.8. Wells were loaded with lysate from either 250, 500, or 1000 cells and an electric field of 30 V/cm was applied for 15 min. The results from this experiment are shown in Figure 4.10. Comparing the wells loaded with 250 cells to the empty wells, one observes clear black bands indicating the separation of gDNA and RNA. This result indicates that it is possible to use the glass slide based device to separate small quantities of nucleic acids. While further sequencing validation will be needed before fully confirming the function of the glass slide based device, it is reasonable to expect that this device will behave in a similar manner to the cassette based device.

4.5 Chapter Summary

This chapter reported on the optimization and validation of the Gel-Seq technology. Emphasis was placed on the cassette based devices, which are more straightforward to fabricate

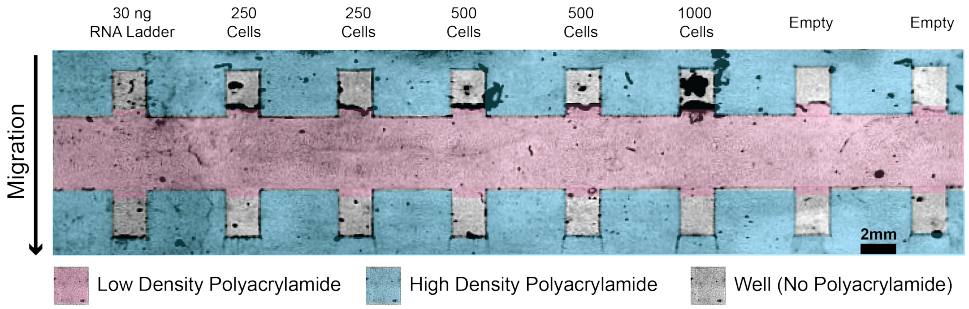


Figure 4.10: Results showing the separation of gDNA and RNA in 250 cells using the glass slide based device.

and therefore a better candidate for widespread adoption. The gel optimization process for the cassette based devices was discussed, followed by results visualizing the successful separation of gDNA and RNA/cDNA using fluorescent imaging. A protocol for recovering the nucleic acids from the devices, known as the crush and soak method, was introduced. This was followed by a discussion of the approaches for generating sequencing libraries from the recovered gDNA and cDNA. In order to validate device performance, sequencing results from the Gel-Seq protocol were compared to tube controls. The results for both the genome and transcriptome processed using the Gel-Seq device correlated well to tube controls. This evidence supports the conclusion that the Gel-Seq device can be used to generate correlated libraries allowing for the direct comparison between the genome and transcriptome. Finally, the focus was turned to the validation of the glass slide based devices. The optimization process of these devices was discussed and results were shown indicating the feasibility of separating gDNA and RNA at the 250 cell level.

Chapter 5

Conclusions and Future Work

This dissertation has discussed the development of Gel-Seq, a new protocol for sequencing both the genome and transcriptome from rare samples. This final chapter summarizes the findings from this dissertation and discusses areas for future research.

5.1 Conclusions

The development of the Gel-Seq protocol opens news doors for genetics researchers. Rather than investigating either the genome or transcriptome from a given sample, using Gel-Seq biologists will now be able to develop correlated data sets for small cell populations. This achievement makes it possible to examine the ways in which changes in DNA impact RNA expression.

This dissertation has chronicled the development of this protocol from its conception through its validation. At the heart of the Gel-Seq protocol is the electrophoretic separation of DNA and RNA based on size; genomic DNA is millions of basepairs long while RNA is only a few thousand nucleotides. Understanding this size difference, two membranes were developed that could be used to separate DNA from RNA. The first membrane, a low density polyacrylamide gel, allows RNA and cDNA molecules to pass through but stops larger genomic DNA. The second membrane, a high density polyacrylamide gel, traps the RNA molecules. Both membranes allow small buffer ions to pass through unimpeded, a necessary condition for electrophoresis.

Two devices were developed as a part of the Gel-Seq protocol. The first device, which can be fabricated using standard lab equipment in an off the shelf gel electrophoresis cassette, is designed for widespread adoption by the biology community. These devices could be made by anyone familiar with standard gel electrophoresis and can separate DNA and RNA from as few as 400 cells. The second device, based on a glass slide form factor, has been designed for separating DNA and RNA in even lower cell counts. In order to fabricate this device, a new polyacrylamide molding process was developed using SU-8 molds and photo-initiated polymerizations. These devices contain polyacrylamide films 250 μm thick. Minimizing the

thickness should facilitate higher recovery yields, allowing for separations below even the 250 cell level reported on in this dissertation.

In parallel to the device development, a biological protocol was designed that was compatible with a physical separation of DNA and RNA. This protocol combines both DNA and RNA library preparation techniques. Recognizing the susceptibility of RNA to degradation, RNA is converted to cDNA before it is separated from gDNA. Once gDNA and cDNA have been separated, the gDNA can be concentrated and then converted to a genomic library using NexteraXT. In parallel, the cDNA sample is amplified with PCR and then converted to a transcriptomic library using NexteraXT. In order to validate both the devices and biological protocol, a stable reference (PC3) was prepared using standard protocols in tube and compared to the Gel-Seq protocol. The results showed high correlation between for both the DNA (R = 0.88) and RNA (R = 0.96) derived libraries. This supports the conclusion that the device can be used to produce correlated genome and transcriptome libraries.

This dissertation has demonstrated the feasibility of the Gel-Seq protocol, however establishing a new protocol is just the beginning. There are many opportunities to apply, refine, and further improve the foundation set forth in these pages. The path forward for this technology will depend on the choices made by future researchers. The next, and final, section lays out several areas for future research.

5.2 Future Work

This section identifies three research areas for future work. The first discusses applying Gel-Seq to biologically interesting samples. The second explores the options for improving and further validating the glass slide based devices. The last addresses integrating alternative RNA-seq protocols within the overall Gel-Seq framework.

Application to Biologically Relevant Samples

This dissertation used sequencing data to demonstrate that that the Gel-Seq protocol could be used to generate matched genome and transcriptome data. This was achieved by using a stable cell line, PC3, which has uniform expression of both the genome and transcriptome across cells. While this was the ideal candidate for technology validation, this sample has no biological relevance.

The development of the Gel-Seq protocol was predicated on a the desire to generate a matched data set for the genome and transcriptome. Recovering data for both the genome and transcriptome makes it possible to investigate the connection between these two data sets. There are a wide range of applications where this type of matched data set could be useful.

As previously mentioned, one application is to investigate the impact of DNA copy number variations on RNA expression. Many types of cancer are driven by the addition or subtraction of large sections (10,000+ base pairs) of the genome. These genomic changes

impact the expression of RNA. Research has already shown that there are a wide range of ways that RNA expression is regulated, including non-coding micro RNAs [48]. Using the Gel-Seq protocol would allow researchers to directly investigate the ways in which copy number variations impact RNA expression.

Another application is to study the spatial relationships of the genome and transcriptome within a particular tissue. Cancerous tumors, for example, do not contain a homogenous genomic profile. Instead tumors are made up of multiple cell types with different mutations [49]. Understanding this heterogeneity is important in discovering new ways to treat cancer. For instance, some chemotherapies will initially appear to arrest tumor growth but after a few weeks the tumors develop resistance to the treatment and continue to grow [50]. Using the Gel-Seq protocol, it becomes possible to sample different regions of the tumor and explore the ways in which this heterogeneity facilitates the development of chemotherapy resistance for both the genome and transcriptome. Such an approach is not currently possible as even adjacent samples may have different genomes or transcriptomes.

While the two applications mentioned here are both related to cancer, this technology is broadly applicable across other areas of biology. The ability to gather information about both the genome and transcriptome at the same time from low input has benefits for geneticists in multiple areas of study due to the scarcity of biologically interesting samples. For instance, researchers studying diseases like Alzheimer's could benefit from the ability to sequence both the genome and transcriptome from neuronal nuclei.

Further Optimization and Validation of the Glass Slide Based Device

As mentioned in Section 4.4, the glass slide based device has only been validated using optical fluorescence measurements. While the cassette based device has demonstrated the feasibility of the Gel-Seq protocol, there is more that can be done to investigate the limits of this technology. The experiments discussed in this dissertation suggest that the glass slide based device can separate DNA and RNA down to the optical limit of detection, however there is no physical reason why the separations cannot be achieved at even smaller quantities.

The next step for this technology is to undergo the same rigorous, sequencing based validation that was performed for the cassette based devices. Once the technology is validated, then the device can be optimized to further minimize the required sample input. This will involve both revising the device geometry as well as optimizing the sample recovery protocol. In order to validate separations below 100 cells, it will also be necessary to modify the Gel-Seq protocol to include a uniform amplification of the genome before tagmentation. Longer term, should the demand arise, this device could be adapted to a high throughput form factor. It is simple to imagine a microfluidic platform that simplifies the loading, separation, and recovery stages.

Adaptation of Gel-Seq to Other RNA-Seq Protocols

One last area warranting further investigation is the adaption of the Gel-Seq protocol for use with other RNA-Seq protocols. As described in Section 3.4, there are a family of protocols that have been developed for generating transcriptome data. Currently Gel-Seq relies on a modification of the Cell Amp protocol to generate transcriptome libraries, however other protocols could also be integrated into the Gel-Seq procedure. For instance, the Smart-Seq protocol is currently the gold standard for generating transcriptome libraries [17]. Recall that it was not selected for the development phase of Gel-Seq due to its high per-reaction cost. Now that the technology has been validated, however, it becomes feasible to think about integrating Smart-Seq into the overall Gel-Seq protocol.

While the protocols are similar, some modifications would be required to integrate Smart-Seq. For example, Smart-Seq produces full length mRNA transcripts, while Cell Amp produced transcripts with a uniform length of $\tilde{1}000$ nucleotides. Full length transcripts may be unable to move through the separation membrane, requiring tweaking of the membrane chemistry or electrophoretic conditions. Due to these variations in the protocol, it will be important to validate the modified protocol by comparing it to tube controls.

One of the beauties of the Gel-Seq protocol is that as new biological protocols are developed, they can be integrated into the Gel-Seq framework. The device provides a platform for the physical separation of DNA and RNA based on size, a property that is consistent regardless of the biological protocol. As researchers continue to push the boundaries of genetics, Gel-Seq can continue to grow and evolve.

Bibliography

[1] Richard Redon et al. "Global variation in copy number in the human genome". In: *Nature* 444.7118 (Nov. 2006), pp. 444–454.

[2] N Brouwers et al. "Alzheimer risk associated with a copy number variation in the complement receptor 1 increasing C3b/C4b binding sites". In: *Molecular psychiatry* 17.2 (2012), pp. 223–233.

[3] Rameen Beroukhim et al. "The landscape of somatic copy-number alteration across human cancers". In: *Nature* 463.7283 (Feb. 2010), pp. 899–905.

[4] Marc A Williams. "Stabilizing the code–methods to preserve RNA prove their worth". In: *Biomarker insights* 5 (2010), p. 139.

[5] H Lodish, A Berk, and SL Zipursky. *Molecular Cell Biology. 4th edition.* New York: W. H. Freeman, 2000. URL: `http://www.ncbi.nlm.nih.gov/books/NBK21729/`.

[6] Yutaka Suzuki et al. "Statistical Analysis of the 5 Untranslated Region of Human mRNA Using Oligo-Capped cDNA Libraries". In: *Genomics* 64.3 (Mar. 2000), pp. 286–297.

[7] Genome Reference Consortium. *Assembly Statistics for GRCh38.p6*. Dec. 2015. URL: `http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/data/`.

[8] Kris A Wetterstrand. *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*. Jan. 2016. URL: `http://www.genome.gov/sequencingcosts`.

[9] Erika Check Hayden et al. "The $1,000 genome". In: *Nature* 507.7492 (2014), pp. 294–295.

[10] Regina M Santella. "Approaches to DNA/RNA extraction and whole genome amplification". In: *Cancer Epidemiology Biomarkers & Prevention* 15.9 (2006), pp. 1585–1587.

[11] Illumina. *Nextera XT DNA Library Preparation Guide*. Part 15031942 Rev. E. Jan. 2015.

[12] Roger S Lasken and Michael Egholm. "Whole genome amplification: abundant supplies of DNA from precious samples or clinical specimens". In: *Trends in Biotechnology* 21.12 (Dec. 2003), pp. 531–535.

[13] Frank B Dean et al. "Comprehensive human genome amplification using multiple displacement amplification". In: *Proceedings of the National Academy of Sciences* 99.8 (2002), pp. 5261–5266.

[14] Jeff Gole et al. "Massively parallel polymerase cloning and genome sequencing of single cells using nanoliter microwells". In: *Nature Biotechnology* 31.12 (Nov. 2013), pp. 1126–1132.

[15] Yohei Sasagawa et al. "Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity". In: *Genome Biology* 14.4 (2013), R31.

[16] Kazuki Kurimoto et al. "An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis". In: *Nucleic Acids Research* 34.5 (Mar. 2006), e42–e42.

[17] Daniel Ramskold et al. "Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells". In: *Nature Biotechnology* 30.8 (July 2012), pp. 777–782.

[18] Michael L Metzker. "Sequencing technologies — the next generation". In: *Nature Reviews Genetics* 11.1 (Dec. 2009), pp. 31–46.

[19] Illumina. *An Introduction to Next-Generation Sequencing Technology*. Jan. 2016. URL: `www.illumina.com/technology/next-generation-sequencing.html`.

[20] Siddharth S Dey et al. "Integrated genome and transcriptome sequencing of the same cell". In: *Nature Biotechnology* (Jan. 2015), pp. 1–7.

[21] Iain C Macaulay et al. "G&T-seq: parallel sequencing of single-cell genomes and transcriptomes". In: *Nature Methods* 12.6 (Apr. 2015), pp. 519–522.

[22] Ajaykumar Gopal et al. "Visualizing large RNA molecules in solution". In: *Rna* 18.2 (2012), pp. 284–299.

[23] Jean-Louis Viovy. "Electrophoresis of DNA and other polyelectrolytes: physical mechanisms". In: *Reviews of Modern Physics* 72.3 (2000), p. 813.

[24] Ronald F Probstein. *Physicochemical hydrodynamics: an introduction*. John Wiley & Sons, 2005.

[25] Joseph Sambrook, EF Fritsch, and T Maniatis. "Agarose gel electrophoresis". In: *Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York* 6 (1989).

[26] Lonza. *Sourcebook for Electrophoresis - Section VII: Separation of DNA in Polyacrylamide Gels*. Mar. 2016. URL: `http://bio.lonza.com/uploads/tx_mwaxmarketingmaterial/Lonza_BenchGuides_SourceBook_Section_VII_-_Separation_of_DNA_in_Polyacrylamide_Gels.pdf`.

[27] Reiner Westermeier. *Electrophoresis in practice*. John Wiley & Sons, 2006.

[28] Jonathan R Brody and Scott E Kern. "History and principles of conductive media for standard DNA electrophoresis". In: *Analytical biochemistry* 333.1 (2004), pp. 1–13.

[29] National Diagnostics. *Fundamentals of Electrophoresis: Electrophoresis Buffers - Choosing the Right Buffer*. Apr. 2016. URL: `https://www.nationaldiagnostics.com/electrophoresis/article/electrophoresis-buffers-choosing-right-buffer`.

[30] Joachim Kohn. "A cellulose acetate supporting medium for zone electrophoresis". In: *Clinica Chimica Acta* 2.4 (1957), pp. 297–303.

[31] Jongyoon Han, Jianping Fu, and Reto B Schoch. "Molecular sieving using nanofilters: Past, present and future". In: *Lab on a Chip* 8.1 (2007), p. 23.

[32] Nancy C Stellwagen. "Electrophoresis of DNA in agarose gels, polyacrylamide gels and in free solution". In: *ELECTROPHORESIS* 30.S1 (June 2009), S188–S195.

[33] Nancy C Stellwagen. "Apparent pore size of polyacrylamide gels: Comparison of gels cast and run in Tris-acetate-EDTA and Tris-borate-EDTA buffers". In: *Electrophoresis* 19.10 (1998), pp. 1542–1547.

[34] Paul Menter. "Acrylamide Polymerization - A Practical Approach". In: *Bio-Rad Tech Note* 1156 (2000).

[35] Todd A Duncombe and Amy E Herr. "Photopatterned free-standing polyacrylamide gels for microfluidic protein electrophoresis". In: *Lab on a Chip* 13.11 (2013), p. 2115.

[36] Robert Lucito et al. "Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation". In: *Genome research* 13.10 (2003), pp. 2291–2305.

[37] Jonathan Sebat et al. "Large-scale copy number polymorphism in the human genome". In: *Science* 305.5683 (2004), pp. 525–528.

[38] Guia Guffanti et al. "Increased CNV-region deletions in mild cognitive impairment (MCI) and Alzheimer's disease (AD) subjects in the ADNI sample". In: *Genomics* 102.2 (2013), pp. 112–122.

[39] Joseph T Glessner et al. "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes". In: *Nature* 459.7246 (2009), pp. 569–573.

[40] Sarah Lamble et al. "Improved workflows for high throughput library preparation using the transposome-based nextera system". In: *BMC biotechnology* 13.1 (2013), p. 104.

[41] Andrew Adey and Jay Shendure. "Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing". In: *Genome research* 22.6 (2012), pp. 1139–1143.

[42] Zhong Wang, Mark Gerstein, and Michael Snyder. "RNA-Seq: a revolutionary tool for transcriptomics". In: *Nature Reviews Genetics* 10.1 (2009), pp. 57–63.

[43] Marc Sultan et al. "A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods". In: *Biochemical and biophysical research communications* 422.4 (2012), pp. 643–646.

[44] Biji T Kurien and R Hal Scofield. "Extraction of Nucleic Acid Fragments from Gels". In: *Analytical biochemistry* 302.1 (Mar. 2002), pp. 1–9.

[45] Joseph Sambrook and David W Russell. "Isolation of DNA fragments from polyacrylamide gels by the crush and soak method". In: *Cold Spring Harb Protoc* (2006).

[46] Jay Devore. *Probability and Statistics for Engineering and the Sciences*. Cengage Learning, 2004.

[47] Alicia Oshlack, Matthew J Wakefield, et al. "Transcript length bias in RNA-seq data confounds systems biology". In: *Biol Direct* 4.1 (2009), p. 14.

[48] Lin He and Gregory J Hannon. "MicroRNAs: small RNAs with a big role in gene regulation". In: *Nature Reviews Genetics* 5.7 (2004), pp. 522–531.

[49] Lawrence A Loeb, Keith R Loeb, and Jon P Anderson. "Multiple mutations and cancer". In: *Proceedings of the National Academy of Sciences* 100.3 (2003), pp. 776–781.

[50] Yves Pommier et al. "Apoptosis defects and chemotherapy resistance: molecular interaction maps and networks". In: *Oncogene* 23.16 (2004), pp. 2934–2949.

[51] MicroChem. *SU-8 2000 Permanent Epoxy Negative Photoresist PROCESSING GUIDELINES FOR: SU-8 2100 and SU-8 2150*. Apr. 2016. URL: http://www.microchem.com/pdf/SU-82000DataSheet2100and2150Ver5.pdf.

[52] National Diagnostics. *Glass Bond / Glass Free For Electrophoretic Glass Support Plates*. Apr. 2016. URL: https://www.nationaldiagnostics.com/sites/default/files/EC620-621_0.pdf.

[53] Lonza. *Gel Slick Glass Plate Coating, Document 18118-0807-4*. Apr. 2016. URL: http://bio.lonza.com/uploads/tx_mwaxmarketingmaterial/Lonza_ManualsProductInstructions_Gel_Slick_Glass_Plate_Coating_-_Protocol.pdf.

[54] Henrik Garoff and Wilhelm Ansorge. "Improvements of DNA sequencing gels". In: *Analytical biochemistry* 115.2 (1981), pp. 450–457.

[55] Takara. *CellAmp Whole Transcriptome Amplification Kit (Real Time) Ver.2 v1204Da*. Apr. 2016. URL: http://www.clontech.com/US/Products/Real-Time_qPCR_and_Reverse_Transcription/RNA_Extraction_and_Protein_Analysis/ibcGetAttachment.jsp?cItemId=55666&fileId=6451413&sitex=10020:22372:US.

Appendix A

Device Recipes

Recipes for the polymer precursors used in fabricating the devices in this dissertation are presented here. To simplify fabrication, two stock solutions were created that were then diluted to the desired concentrations: 50%T, 5%C and 40%T, 3%C.

A.1 Glass Slide Devices

High Density Membrane		
Starting Acryalmide Solution	50%	T
	5%	C
Target Solution	**40%**	**T**
	5%	**C**
Total Volume	**10**	**mL**
Acrylamide (19:1)	8	mL
Water	2	mL
VA-086 (Powder)	100	mg

Low Density Membrane		
Starting Acryalmide Solution	40%	T
	3%	C
Target Solution	**8%**	**T**
	3%	**C**
Total Volume	**10**	**mL**
Acrlyamide (29:1)	2	mL
Water	8	mL
VA-086 (Powder)	100	mg

A.2 Cassette Based Devices

Filler Gel		
Starting Acryalmide Solution	40%	T
	3%	C
Target Solution	**4.00%**	**T**
	3%	**C**
Total Volume	**28**	**mL**
d H2O	17.9	mL
Sucrose (50%)	4.5	mL
10X TBE	2.8	mL
Acrylamide Stock (29:1)	2.8	mL
APS (10%)	182	µL
Temed	10.5	µL

High Density Membrane		
Starting Acryalmide Solution	50%	T
	5%	C
Target Solution	**40%**	**T**
	5%	**C**
Total Volume	**2**	**mL**
Water	0.08	mL
Sucrose (50%)	0.32	mL
Acrylamide (19:1)	1.6	mL
APS (10%)	25	µL
Temed	0.5	µL

Low Density Membrane		
Starting Acryalmide Solution	40%	T
	3%	C
Target Solution	**4%**	**T**
	3%	**C**
Total Volume	**10**	**mL**
d H2O	8	mL
10X TBE	1	mL
Acrylamide (29:1)	1	mL
APS	65	µL
Temed	3.75	µL

This makes enough for 4 cassetes, scale as necessary.

Appendix B

Fabrication Protocol for Glass Slide Devices

While the glass slide based devices themselves are relatively straightforward to fabricate, they require the creation of an SU-8 mold and surface treatment of the glass slide substrate. This Appendix covers the fabrication of the SU-8 mold, surface treatments of both the mold and the substrate, and the fabrication of the device.

B.1 SU-8 Mold Fabrication

This protocol is based loosely on the data sheet for SU-8 2100 provided by MicroChem [51]. There are, however, substantial changes that were discovered during experimental optimization of the process.

1. Clean a glass wafer with oxygen plasma, 5 min at 200 mW/cm^2.
2. Spin coat a 250 μm layer of SU-8 2100 using the following parameters:
   a) Place wafer into spin coater.
   b) Pour 25 mL SU-8 into center of wafer.
   c) Spin at 500 RPM for 30 seconds, acceleration of 100 RPM/sec.
   d) Spin at 100 RPM for 30 seconds, acceleration of 300 RPM/sec.
   e) Soft Bake at 110°C for 60 min.
3. Using a photomask with the desired geometry, expose the wafer to 540 mJ/cm^2 of UV energy.
4. Post exposure bake: place wafer on a hotplate at 110°C for a of 16 minutes. The features will become visible as the photoresist cures.

5. While the wafer is baking, pour SU-8 developer into a beaker and place the beaker into a sonicator.
6. Develop the wafer by submerging it in a SU-8 developer and sonicating for 4 to 5 minutes. Check constantly to ensure the wafer is not overdeveloped.
7. Once the wafer is fully developed, wash with isopropyl alcohol and then dry with filtered, pressurized air.
8. Measure the features height with a surface profilometer to confirm SU-8 thickness.

B.2 SU-8 Mold Silane Treatment

This protocol developed by combining protocols from product manuals for Glass Free [52] and Gel Slick [53].

1. Make sure SU-8 coated wafers are clean and dry before starting this process.
2. Make a 5% solution of Dichlorodimethylsilane (DCDMS) in Toluene (e.g. 5ml DCDMS in 95ml toluene).
3. Immerse wafers for 5 minutes into the solution.
4. Rinse wafers with Toluene.
5. Rinse wafers with Methanol.
6. Dry wafers with compressed air.
7. Molds can be stored indefinitely after this first treatment if stored in a clean wafer box and protected from light.
8. Just before using a mold to cast a device, apply 2.0 ml of Gel Slick solution to the wafer.
9. Using a silk lint free cloth, wipe the solution in a circular motion to evenly distribute it across the surface of the mold.
10. Allow the solution to dry completely to a faint white haze, approximately 5 minutes.
11. Polish the surface using a clean silk, lint free cloth to remove the haze.
12. Cast a device using the mold immediately and then reapply Gel Slick solution before the mold is used again.

B.3 Glass Slide Adhesive Silane Treatment

This protocol is adapted from Garoff [54].

1. Clean the desired number of glass slides in soap and rinse thoroughly with deionized water.
2. Place slides on 100°C hot plate and allow to dry fully (At least 10 min).
3. Prepare the bonding solution by following the steps below:
   a) In an erlenmeyer flask, dilute 1 mL of 3-(trimethoxysilyl)propyl methacrylate in 200 mL of ethanol .
   b) Mix solution thoroughly to combine.
   c) Just before use, add 6 mL of dilute acetic acid (1:10 glacial acetic acid:water).
4. Place glass slides in a large pyrex dish and pour the bonding solution over the slides.
5. Allow the plates to react for 3 minutes.
6. Remove plates from solution, rinse with ethanol, and dry with compressed air.
7. Store plates away from light in a slide holder until use.

B.4 Glass Slide Device Fabrication

This protocol was inspired from work by Duncombbe [35], though it has been substantially changed from his original protocol.

1. Mix the precursor solutions as outlined in Appendix A.1, protect them from light as much as possible.
2. Place the treated SU-8 mold face up into a pyrex petri dish.
3. Place four latex spacers on the mold (typically 250 μm).
4. Place the glass slide, treated side down, on top of the latex spacers to create the mold cavity for the high density region of the gel.
5. Using a disposable bulb pipette, fill the mold cavity with the high density gel precursor. (Some experimentation will be necessary to determine the best approach for eliminating air bubbles based on the device geometry.)
6. Configure a photomask aligner for flood exposure.
7. Place the entire assembly into the aligner, ensure there is clearance for the petri dish.

8. Expose the assembly to UV light (13 mW/cm^2) for 30 - 60 seconds depending on desired device properties.

9. Separate the glass slide from the SU-8 mold using a pair of wafer tweezers.

10. Place the partially fabricated device on top of the second SU-8 mold to create the mold cavity for the low density region.

11. Using a disposable bulb pipette, fill the mold cavity with low density gel precursor. (Some experimentation will be necessary to determine the best approach for eliminating air bubbles based on the device geometry.)

12. Once again, place the entire assembly into the aligner.

13. Expose the assembly to UV light (13 mW/cm^2) for 10 - 30 seconds depending on desired device properties.

14. Separate the finished device from the SU-8 mold using a pair of wafer tweezers.

15. Store the device in TBE buffer until use.

Appendix C

Fabrication Protocol for Cassette Based Devices

The fabrication protocol for the cassette based devices follows a similar approach to the fabrication of stacking gels. Stacking gels consist of two regions, a small section gel to concentrate the sample and a large section for sample separation. The devices reported on in this dissertation contain three layers, the filler gel, high density membrane, and low density membrane. A detailed protocol is as follows:

C.1 Prepare Precursors

1. Combine all of the precursors for the three layers listed in the recipe in Appendix A.2, but do not add the TEMED or APS.
2. Vortex the precursors to mix thoroughly.
3. Remove the air from the solution to reduce polymerization time. This can be accomplished by poking a hole in the top of the tubes using a needle. The needle can then be connected to house vacuum and the tubes submerged into a sonicator. Wait 3-5 minutes per sample, until air bubbles no longer are being drawn out of the solution.
4. Degass 50 mL of deionized water in a similar fashion as the polymer precursors.

C.2 Cast Device Layers

1. Add the appropriate amount of APS and TEMED to the filler gel precursor.
2. Vortex the sample to mix thoroughly.
3. Using a pipette, add 6 mL of the filler gel precursor to the gel cassette.

4. Using a pipette, fill the remainder of the cassette with the deionized, degassed water.
5. Allow the device to sit upright at room temperature undisturbed for at least one hour, up to overnight. This allows for the full polymerization of the filler gel.
6. Pour off the water overlay and dry the interface carefully using a gentle stream of compressed air.
7. Add the appropriate amount of APS and TEMED to the high density gel precursor.
8. Working quickly as the high density gel polymerizes rapidly, vortex the sample to mix and add 0.350 mL to the cassette on top of the filler gel.
9. Rotate the cassette from side to side to ensure the precursor solution is evenly distributed across the interface.
10. Carefully pipette water to fill the remainder of the cassette. Add the water from the center of the cassette opening to maintain a uniform high density layer.
11. Allow the device to sit upright at room temperature undisturbed for at least 15 minutes hour, up to several hours. The high density layer polymerizes more quickly than the filler gel.
12. Pour off the water overlay and dry the interface carefully using a gentle stream of compressed air.
13. Add the appropriate amount of APS and TEMED to the low density gel precursor.
14. Vortex the sample to mix and then fill the remainder of the cassette with the low density precursor.
15. Insert the gel comb into the cassette, this will be used to define the loading wells.
16. Pipette additional precursor around the comb to reduce problems with gel formation around the comb / gel interface.
17. Allow the device to polymerize for several hours, ideally overnight, to ensure all layers have fully polymerized.

Appendix D

Detailed Protocol for Cassette Based Gel-Seq

This protocol has been derived in part from the Cell Amp protocol [55].

D.1 Sample Preparation

1. Prepare cell suspension with 1 - 1000 cells/μL
2. Combine the following reagents to create a master mix, scale up for multiple reactions. (Use best practices, do not pipette less than 0.5 μL, mix reagents by pipetting up and down 5 times.)

Reagent	Volume (µL)
Lysis Buffer (Cell Amp Kit) (0.18 M KCl, 36 mM Tris-HCL pH 8.3, 1.8% NP-40, 18 mM DTT)	1.25
Recombinant RNase Inhibitor (40 U/µL)	0.25
RT dT Primer 2 (From Cell Amp kit)	0.1
dNTP Mixture (2.5 mM each)	0.1
RNase Free H2O	2.1
Total	3.8

3. Pipette 4.8 μL of the master mix into the desired number of reaction tubes
4. Add 1 μL of the cell suspension to each tube.

5. Mix the solution by pipetting up and down 5 times.
6. Incubate at 70°C for 90 seconds to lyse the cells.
7. Place the samples on ice.
8. Create a second mastermix for reverse transcription by scaling up the reagents below.

Reagent	Volume (µL)
$MgCl_2$ (Cell Amp Kit)	0.3
RT Enzyme Mix (Cell Amp Kit)	0.3
RNase Free H2O	0.2
Total	0.8

9. Add 0.8μL of the second master mix to each reaction, bringing the total volume to 5.6 μL. Mix well.
10. Synthesize cDNA by incubating the tubes at 42°C for 5 minutes and then stop the reaction by incubating at 85°C for 10 seconds.
11. Add 0.6 μL Exonuclease I to each tube, resulting in a final volume of 6.2 μL. Incubate for 15 minutes at 37°C to digest the single stranded DNA primers and then stop the reaction by incubating for 15 minutes at 80°C.
12. Prepare a third master mix as follows:

Reagent	Volume (µL)
TdT Buffer (Cell Amp Kit)	1.2
dATP (90 mM)	0.2
TdT Enzyme Mix (Cell Amp Kit)	0.45
RNase Free dH2O	4.15
Total	6

13. Add 6 μL of mastermix to each tube and mix well, for a total volume of 12.2 μL.

14. Perform the Poly-A tailing reaction by incubating at 37°C for 15 minutes followed by an incubation at 70°C for 10 minutes to stop the reaction.
15. Remove the samples from the incubator and place on ice.
16. Add 12.5 μL of DNA loading dye to each sample, bringing the total volume to $\tilde{2}5$ μL.

D.2 Electrophoretic Separation

1. Place the Gel-Seq cassette into the gel box and attach to the power supply.
2. Apply 200 volts across the device for 20 minutes to warm the device.
3. Turn the power supply off.
4. Using a pipette, add one sample to each lane of the device.
5. Turn the power supply on and allow the samples to undergo electrophoretic separation for 60 minutes.
6. Turn off the power power supply and remove the cassette from the gel box.

D.3 Imaging

1. Carefully open the gel cassette using a paint spatula to separate the two plastic halves. The gel will adhere to one of the two halves of the cassette.
2. Cut away the filler gel and discard.
3. Prepare 30 mL of 1x Sybr gold solution in the top of a pipette box.
4. Using tweezers or a scalpel, carefully peel the high and low density regions of the device off of the opened cassette and place it into the Sybr Gold stain solution.
5. Cover the sybr gold solution to protect it from light and shake the container gently for 5 minutes.
6. Place plastic wrap (such as saran wrap) into the imaging tool. This will make it easier to transport the gel for future steps.
7. Remove the gel from the stain solution and place it onto the plastic wrap.
8. Image the device to record the locations of the gDNA and cDNA.

D.4 Sample Recovery

1. Move the plastic wrap from the imaging tool to the a fluorescent box used for cutting gels.
2. Wearing the protective glasses, use the fluorescent box to visualize the location of the DNA within the device. While not necessary, this can be helpful in the cutting process.
3. Using a scalpel, cut out a section of gel containing DNA. Cut the smallest amount of gel necessary to capture the sample to maximize sample recovery.
4. Using tweezers, place the gel section into a clean strip tube.
5. Between each sample wipe the tools with ethanol to minimize the possibility of cross contamination.
6. Repeat this process until all of the samples have been cut out from the gel.
7. Use a pipette tip to crush the gel contained within one of the strip tubes. (The mechanics of this are similar to using a mortar and pestle.)
8. Once the sample has been crushed, pipette 20 - 80 μL of clean water into the tube and slowly remove the pipette tip used to crush the sample. (Adding the water before removing the pipette tip reduces the amount of gel that sticks to the pipette tip.)
9. Repeat these steps until all of the samples have been crushed. Remember to use a new pipette tip for each sample.
10. Place the tubes into an incubator at 37°C and shake on a vortex mixer for 8 - 12 hours.
11. Remove the tubes from the incubator and spin in a centrifuge to collect the gel at the bottom of the tubes.
12. Remove the supernatant from each tube and place into a new, clean strip tube.
13. Discard the tubes containing the gel.

D.5 Library Preparation

DNA Library Preparation

This protocol closely follows the standard Nextera XT protocol [11], however the volumes have been reduced by half. The protocol requires the following Nextera XT reagents: Tagment DNA buffer (TD buffer), Amplicon Tagment Mix (ATM), Neutralize Tagment Buffer (NT), Nextera PCR Master Mix (NPM), and index primers.

1. Add 1 μL protease and incubate at 50°C for 15 min and then 70°C for 15 min. (This makes the DNA accessible for library preparation.)
2. Using a vacufuge, reduce the sample volume for the gDNA samples from 20 μL to 5 μL. Depending on the model and number of samples this can take 10 - 40 minutes.
3. Quantify one sample using qubit to confirm the sample concentration is roughly 0.2 ng/μL. Dilute or further concentrate the samples as necessary.
4. Combine 5 μL of TD buffer with 2.5 μL of sample. Mix well.
5. Add 2.5 μl of ATM to each tube, incubate at 55°C for 5 minutes and then hold at 10°C.
6. As soon as the sample reaches 10°C remove samples and add 2.5 μL NT Buffer to each tube. Mix well.
7. Leave the plate at room temperature for 5 minutes.
8. Add 7.5 μL NPM to each tube.
9. Add 2.5 μL index 2 primers, use a different index number for each tube.
10. Add 2.5 μL index 1 primers, use the same index if possible.
11. Mix well by pipetting up and down.
12. Perform PCR using the following program on a thermal cycler:
    a) 72°C for 3 minutes
    b) 95°C for 30 seconds
    c) 12 cycles of:
        i. 95°C for 10 seconds
        ii. 55°C for 30 seconds
        iii. 72°C for 30 seconds
    d) 72°C for 5 minutes
    e) Hold at 10C
13. Use AmpureXP beads or another standard size selection protocol to isolate the relevant fragments for sequencing, typically 200 - 800 basepairs in length.
14. Store libraries at 4° for several or freeze at -20° for longer term storage.

RNA Library Preparation

1. Depending on the starting concentration of RNA, it may be necessary to vacufugue the RNA samples slightly to concentrate the RNA. For example, a 60 μL sample could be concentrated to 15 μL. This will need to be determined on a case by case basis.

2. Prepare a PCR mastermix based on the reagents below. Scale up as necessary depending on the number of reactions.

Reagent	Volume (µL)
PCR Primer Mix 2 (Cell Amp)	0.75
KAPA Sybr Fast qPCR MasterMix (2x)	12.5
Total	13.25

3. Combine 11.75 μL of the sample recovered from the gel with 13.25 μL of the PCR mastermix. Mix well

4. Perform PCR under the following conditions.

   a) 95°C 1 minute
   b) 50°C 1 minute
   c) 72°C 3 minutes
   d) 20 cycles of:
      i. 95°C 30 seconds
      ii. 67°C 1 minute
      iii. 72°C 3 minutes
   e) 72°C for 10 minutes

5. Use Qubit to quantify the post PCR DNA concentration for each sample.

6. Prepare 10 μL aliquots from each sample that are diluted to a concentration of 0.2 ng/μL.

7. Follow the DNA library preparation protocol from step 4.

Appendix E

Detailed Protocol for Glass Slide Based Gel-Seq

The glass slide based Gel-Seq is broken into three steps: test setup, sample preparation, and separation. From start to finish the protocol takes between two and four hours.

E.1 Test Setup Configuration

1. Place electrode wicks into 0.5X TBE buffer and allow to soak for at least 3 minutes.
2. Remove the glass slide device from the storage container.
3. Place the device into the 3D printed holder.
4. Place the electrode wicks on the device and place the graphite electrodes on top of the wicks.
5. Attach the cross bar and secure the wicks to the device using the thumbscrews.
6. Connect the power supply to the test setup.

E.2 Sample Preparation

1. Dilute the biological sample to the desired cell concentration with PBS, typically 100 - 1000 cells/μL. This will be further diluted before loading.
2. Lyse the cells in tube by combining 1 part sample with 1.5 parts cell lysis buffer (final concentration 50 mM KCL, 10 mM Tris-HCl, 0.5 % NP-40, 5 mM DTT, and 10 units RNAse inhibitor).
3. To ensure thorough lysis, incubate the cells at 70°C for 90 seconds.

4. Store the sample on ice until loading it into the device.

E.3 Separation

1. Carefully use the pipette to remove the buffer from the loading wells.
2. Pipette the sample up and down 5 time to mix thoroughly.
3. Pipette 1 μL of sample into each loading well of the device.
4. Once all loading wells have been filled, carefully pipette mineral oil over the entire device to keep the buffer from evaporating during separation.
5. Turn on the power supply to apply the desired voltage across the device. (Typically 30 V/cm)
6. Allow the samples to migrate for 20 minutes.
7. After migration, place the slide into a solution of 1X Sybr Gold. Cover the container to protect if from light and gently shake for 5 to 30 minutes.
8. Remove the device from the solution and place it inside a gel doc or other fluorescent imaging tool. Adjust the exposure time to obtain the best image quality.